Haruo Kubozono (Ed.)

Handbook of Japanese Phonetics and Phonology

Handbooks of Japanese Language and Linguistics

Edited by
Masayoshi Shibatani
Taro Kageyama

Volume 2

Handbook of Japanese Phonetics and Phonology

Edited by
Haruo Kubozono

DE GRUYTER
MOUTON

ISBN 978-1-61451-252-3
e-ISBN (PDF) 978-1-61451-198-4
e-ISBN (EPUB) 978-1-5015-0059-6
ISSN 2199-2851

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: RoyalStandard, Hong Kong
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Preface

The project of compiling a series of comprehensive handbooks covering major fields of Japanese linguistics started in 2011, when Masayoshi Shibatani received a commission to edit such volumes as series editor from De Gruyter Mouton. As the planning progressed, with the volume titles selected and the volume editors assigned, the enormity of the task demanded the addition of a series co-editor. Taro Kageyama, Director-General of the National Institute for Japanese Language and Linguistics (NINJAL), was invited to join the project as a series co-editor. His participation in the project opened the way to make it a joint venture between NINJAL and De Gruyter Mouton. We are pleased to present the *Handbooks of Japanese Language and Linguistics* (*HJLL*) as the first materialization of the agreement of academic cooperation concluded between NINJAL and De Gruyter Mouton.

The HJLL Series is composed of twelve volumes, primarily focusing on Japanese but including volumes on the Ryukyuan and Ainu languages, which are also spoken in Japan, as well as some chapters on Japanese Sign Language in the applied linguistics volume.

- Volume 1: *Handbook of Japanese Historical Linguistics*
- Volume 2: *Handbook of Japanese Phonetics and Phonology*
- Volume 3: *Handbook of Japanese Lexicon and Word Formation*
- Volume 4: *Handbook of Japanese Syntax*
- Volume 5: *Handbook of Japanese Semantics and Pragmatics*
- Volume 6: *Handbook of Japanese Contrastive Linguistics*
- Volume 7: *Handbook of Japanese Dialects*
- Volume 8: *Handbook of Japanese Sociolinguistics*
- Volume 9: *Handbook of Japanese Psycholinguistics*
- Volume 10: *Handbook of Japanese Applied Linguistics*
- Volume 11: *Handbook of the Ryukyuan Languages*
- Volume 12: *Handbook of the Ainu Language*

Surpassing all currently available reference works on Japanese in both scope and depth, the *HJLL* series provides a comprehensive survey of nearly the entire field of Japanese linguistics. Each volume includes a balanced selection of articles contributed by established linguists from Japan as well as from outside Japan and is critically edited by volume editors who are leading researchers in their individual fields. Each article reviews milestone achievements in the field, provides an overview of the state of the art, and points to future directions of research. The twelve titles are thus expected individually and collectively to contribute not only to the enhancement of studies on Japanese on the global level but also to the opening up of new perspectives for general linguistic research from both empirical and theoretical standpoints.

The *HJLL* project has been made possible by the active and substantial participation of numerous people including the volume editors and authors of individual

chapters. We would like to acknowledge with gratitude the generous support, both financial and logistic, given to this project by NINJAL. We are also grateful to John Haig (retired professor of Japanese linguistics, the University of Hawai‘i at Mānoa), serving as copy-editor for the series. In the future, more publications are expected to ensue from the NINJAL-Mouton academic cooperation.

Masayoshi Shibatani, Deedee McMurtry Professor of Humanities and Professor of Linguistics, Rice University/Professor Emeritus, Kobe University
Taro Kageyama, Director-General, National Institute for Japanese Language and Linguistics (NINJAL)/Professor Emeritus, Kwansei Gakuin University

Masayoshi Shibatani and Taro Kageyama

Introduction to the *Handbooks of Japanese Language and Linguistics*

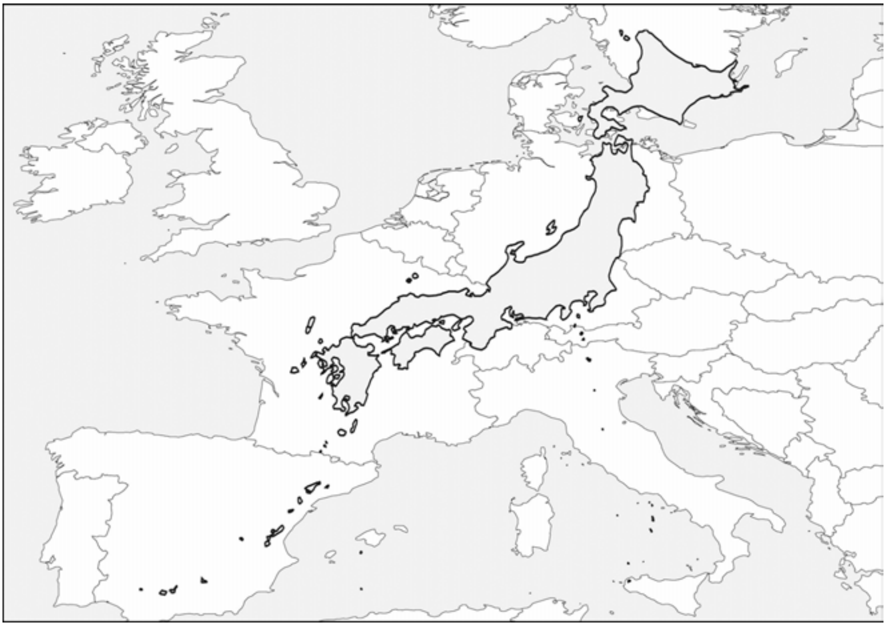
Comprising twelve substantial volumes, the *Handbooks of Japanese Language and Linguistics* (HJLL) series provides a comprehensive survey of practically all the major research areas of Japanese linguistics on an unprecedented scale, together with surveys of the endangered languages spoken in Japan, Ryukyuan and Ainu. What follows are introductions to the individual handbooks, to the general conventions adopted in this series, and the minimum essentials of contemporary Standard Japanese. Fuller descriptions of the languages of Japan, Japanese grammar, and the history of the Japanese language are available in such general references as Martin (1975), Shibatani (1990), and Frellesvig (2010).

1 Geography, Population, and Languages of Japan

Japan is situated in the most populous region of the world – Asia, where roughly one half of the world population of seven billion speak a variety of languages, many of which occupy the top tier of the ranking of the native-speaker population numbers. Japanese is spoken by more than 128 million people (as of 2013), who live mostly in Japan but also in Japanese emigrant communities around the world, most notably Hawaii, Brazil and Peru. In terms of the number of native speakers, Japanese ranks ninth among the world's languages. Due partly to its rich and long literary history, Japanese is one of the most intensely studied languages in the world and has received scrutiny under both the domestic grammatical tradition and those developed outside Japan such as the Chinese philological tradition, European structural linguistics, and generative grammar developed in America. The *Handbooks of Japanese Language and Linguistics* intend to capture the achievements garnered over the years through analyses of a wide variety of phenomena in a variety of theoretical frameworks.

As seen in Map 1, where Japan is shown graphically superimposed on Continental Europe, the Japanese archipelago has a vast latitudinal extension of approximately 3,000 kilometers ranging from the northernmost island, roughly corresponding to Stockholm, Sweden, to the southernmost island, roughly corresponding to Sevilla, Spain.

Contrary to popular assumption, Japanese is not the only language native to Japan. The northernmost and southernmost areas of the Japanese archipelago are inhabited by people whose native languages are arguably distinct from Japanese. The southernmost sea area in Okinawa Prefecture is dotted with numerous small islands



Map 1: *Japan as overlaid on Europe*
Source: Shinji Sanada. 2007. *Hōgen wa kimochi o tsutaeru [Dialects convey your heart].* Tokyo: Iwanami, p. 68.

where Ryukyuan languages are spoken. Until recent years, Japanese scholars tended to treat Ryukyuan language groups as dialects of Japanese based on fairly transparent correspondences in sounds and grammatical categories between mainland Japanese and Ryukyuan, although the two languages are mutually unintelligible. Another reason that Ryukyuan languages have been treated as Japanese dialects is that Ryukyuan islands and Japan form a single nation. In terms of nationhood, however, Ryukyu was an independent kingdom until the beginning of the seventeenth century, when it was forcibly annexed to the feudal domain of Satsuma in southern Kyushu.

A more recent trend is to treat Ryukyuan as forming a branch of its own with the status of a sister language to Japanese, following the earlier proposals by Chamberlain (1895) and Miller (1971). Many scholars specializing in Ryukyuan today even confer language status to different language groups within Ryukyuan, such as Amami language, Okinawan language, Miyako language, etc., which are grammatically distinct to the extent of making them mutually unintelligible. The prevailing view now has Japanese and Ryukyuan forming the Japonic family as daughter languages of Proto-Japonic. HJLL follows this recent trend of recognizing Ryukyuan as a sister language to Japanese and devotes one full volume to it. The ***Handbook of the Ryukyuan Languages*** provides the most up-to-date answers pertaining to Ryukyuan

language structures and use, and the ways in which these languages relate to Ryukyuan society and history. Like all the other handbooks in the series, each chapter delineates the boundaries and the research history of the field it addresses, comprises the most important and representative information on the state of research, and spells out future research desiderata. This volume also includes a comprehensive bibliography of Ryukyuan linguistics.

The situation with Ainu, another language indigenous to Japan, is much less clear as far as its genealogy goes. Various suggestions have been made relating Ainu to Paleo-Asiatic, Ural-Altaic, and Malayo-Polynesian or to such individual languages as Gilyak and Eskimo, besides the obvious candidate of Japanese as its sister language. The general consensus, however, points to the view that Ainu is related to Japanese quite indirectly, if at all, via the Altaic family with its Japanese-Korean subbranch (see Miller 1971; Shibatani 1990: 5–7 for an overview). Because Ainu has had northern Japan as its homeland and because HJLL is also concerned with various aspects of Japanese linguistics scholarship in general, we have decided to include a volume devoted to Ainu in this series. The ***Handbook of the Ainu Language*** outlines the history and current state of the Ainu language, offers a comprehensive survey of Ainu linguistics, describes major Ainu dialects in Hokkaido and Sakhalin, and devotes a full section to studies dealing with typological characteristics of the Ainu language such as polysynthesis and incorporation, person marking, plural verb forms, and aspect and evidentials.

2 History

Japan's rich and long literary history dates back to the seventh century, when the Japanese learned to use Chinese characters in writing Japanese. Because of the availability of abundant philological materials, the history of the Japanese language has been one of the most intensely pursued fields in Japanese linguistics. While several different divisions of Japanese language history have been proposed, Frellesvig (2010) proposes the following four linguistic periods, each embracing the main political epochs in Japanese history.

1.	Old Japanese	700–800	(Nara period, 712–794)
2.	Early Middle Japanese	800–1200	(Heian period, 794–1185)
3.	Late Middle Japanese	1200–1600	(Kamakura period, 1185–1333; Muromachi period, 1333–1573)
4.	Modern Japanese	1600–	(Edo, 1603–1868; Meiji, 1868–1912; Taishō, 1912–1926; Shōwa, 1926–1989; Heisei, 1989–)

This division reflects a major gulf between Pre-modern and Modern Japanese caused by some radical changes in linguistic structure during the Late Middle Japanese period. Modern Japanese is often further subdivided into Early Modern (Edo, 1603–1868), Modern (Meiji, 1868–1912; Taishō, 1912–1926), and Present-day Japanese (Shōwa, 1926–1989; Heisei, 1989–).

The ***Handbook of Japanese Historical Linguistics*** will present the latest research on better studied topics, such as segmental phonology, accent, morphology, and some salient syntactic phenomena such as focus constructions. It will also introduce areas of study that have traditionally been underrepresented, ranging from syntax and Sinico-Japanese (*kanbun*) materials to historical pragmatics, and demonstrate how they contribute to a fuller understanding of the overall history of Japanese, as well as outlining larger-scale tendencies and directions in changes that have taken place within the language over its attested history. Major issues in the reconstruction of prehistoric Japanese and in the individual historical periods from Old Japanese to Modern Japanese are discussed including writing and the materials for historical studies, influences of Sinico-Japanese on Japanese, the histories of different vocabulary strata, the history of honorifics and polite language, generative diachronic syntax, and the development of case marking.

3 Geographic and Social Variations

Because of the wide geographical spread of the Japanese archipelago from north to south, characterized by high mountain ranges, deep valleys, and wide rivers as well as numerous islands, Japanese has developed a multitude of dialects, many of which differ from each other in a way more or less like current descendants of the Romance language family. Like the historical studies, the research tradition of dialect studies has a unique place in Japanese linguistics, which has also attracted a large number of students, amateur collectors of dialect forms as well as professional linguists. The ***Handbook of Japanese Dialects*** surveys the historical backdrop of the theoretical frameworks of contemporary studies in Japanese geolinguistics and includes analyses of prominent research topics in cross-dialectal perspectives, such as accentual systems, honorifics, verbs of giving, and nominalizations. The volume also devotes large space to sketch grammars of dialects from the northern island of Hokkaido to the southern island of Kyushu, allowing a panoramic view of the differences and similarities in the representative dialects throughout Japan.

Besides the physical setting fostering geographic variations, Japanese society has experienced several types of social structure over the years, starting from the time of the nobility and court life of the Old and Early Middle Japanese periods, through the caste structure of the feudalistic Late Middle and Early Modern Japanese periods, to the modern democratic society in the Modern and Present-day Japanese

periods. These different social structures spawned a variety of social dialects including power- and gender-based varieties of Japanese. The ***Handbook of Japanese Sociolinguistics*** examines a wide array of sociolinguistic topics ranging from the history of Japanese sociolinguistics, including foreign influences and internal innovations, to the central topics of variations due to social stratification, gender differences, and discourse genre. Specific topics include honorifics and women's speech, critical discourse analysis, pragmatics of political discourse, contact-induced change, emerging new dialects, Japanese language varieties outside Japan, and language policy.

4 Lexicon and Phonology

The literary history of Japan began with early contacts with China. Chinese apparently began to enrich the Japanese lexicon in even pre-historic periods, when such deeply assimilated words as *uma* 'horse' and *ume* 'plum' are believed to have entered the language. Starting in the middle of the sixth century, when Buddhism reached Japan, Chinese, at different periods and from different dialect regions, has continuously contributed to Japanese in an immeasurable way affecting all aspects of grammar, but most notably the lexicon and the phonological structure, which have sustained further and continuous influences from European languages from the late Edo period on. Through these foreign contacts, Japanese has developed a complex vocabulary system that is composed of four lexical strata, each with unique lexical, phonological, and grammatical properties: native Japanese, mimetic, Sino-Japanese, and foreign (especially English).

The ***Handbook of Japanese Lexicon and Word Formation*** presents a comprehensive survey of the Japanese lexicon, word formation processes, and other lexical matters seen in the four lexical strata of contemporary Japanese. The agglutinative character of the language, coupled with the intricate system of vocabulary strata, makes it possible for compounding, derivation, conversion, and inflection to be closely intertwined with syntactic structure, giving rise to theoretically intriguing interactions of word formation processes and syntax that are not easily found in inflectional, isolating, or polysynthetic types of languages. The theoretically oriented studies associated with these topics are complemented by those oriented toward lexical semantics, which also bring to light theoretically challenging issues involving the morphology-syntax interface.

The four lexical strata characterizing the Japanese lexicon are also relevant to Japanese phonology as each stratum has some characteristic sounds and sound combinations not seen in the other strata. The ***Handbook of Japanese Phonetics and Phonology*** describes and analyzes the basic phonetic and phonological structures of modern Japanese with main focus on standard Tokyo Japanese, relegating the topics of dialect phonetics and phonology to the *Handbook of Japanese Dialects*.

The handbook includes several chapters dealing with phonological processes unique to the Sino-Japanese and foreign strata as well as to the mimetic stratum. Other topics include word tone/accent, mora-timing, sequential voicing (*rendaku*), consonant geminates, vowel devoicing and diphthongs, and the appearance of new consonant phonemes. Also discussed are phonetic and phonological processes within and beyond the word such as rhythm, intonation, and the syntax-phonology interface, as well as issues bearing on other subfields of linguistics such as historical and corpus linguistics, L1 phonology, and L2 research.

5 Syntax and Semantics

Chinese loans have also affected Japanese syntax, though the extent is unclear to which they affected Japanese semantics beyond the level of lexical semantics. In particular, Chinese loans form two distinct lexical categories in Japanese – verbal nouns, forming a subcategory of the noun class, and adjectival nouns (*keiyō dōshi*), which are treated as forming major lexical categories, along with noun, verb, and adjective classes, by those who recognize this as an independent category. The former denote verbal actions, and, unlike regular nouns denoting objects and thing-like entities, they can function as verbs by combining with the light verb *suru*, which is obviously related to the verb *suru* 'do'. The nominal-verbal Janus character of verbal nouns results in two widely observed syntactic patterns that are virtually synonymous in meaning; e.g., *benkyoo-suru* (studying-DO) 'to study' and *benkyoo o suru* (studying ACC do) 'do studying'. As described in the *Handbook of Japanese Lexicon and Word Formation*, the lexical category of adjectival noun has been a perennial problem in the analysis of Japanese parts of speech. The property-concept words, e.g., *kirei* 'pretty', *kenkoo* 'health/healthy', falling in this class do not inflect by themselves unlike native Japanese adjectives and, like nouns, require the inflecting copula *da* in the predication function – hence the label of adjectival noun for this class. However, many of them cannot head noun phrases – the hallmark of the nominal class – and some of them even yield nouns via *-sa* nominalization, which is not possible with regular nouns.

The Lexicon-Word Formation handbook and the ***Handbook of Japanese Syntax*** make up twin volumes because many chapters in the former deal with syntactic phenomena, as the brief discussion above on the two Sino-Japanese lexical categories clearly indicates. The syntax handbook covers a vast landscape of Japanese syntax from three theoretical perspectives: (1) traditional Japanese grammar, known as *kokugogaku* (lit. national-language study), (2) the functional approach, and (3) the generative grammar framework. Broad issues analyzed include sentence types and their interactions with grammatical verbal categories, grammatical relations (topic, subject, etc.), transitivity, nominalization, grammaticalization, voice (passives and

causatives), word order (subject, scrambling, numeral quantifier, configurationality), case marking (*ga/no* conversion, morphology and syntax), modification (adjectives, relative clause), and structure and interpretation (modality, negation, prosody, ellipsis). These topics have been pursued vigorously over many years under different theoretical persuasions and have had important roles in the development of general linguistic theory. For example, the long sustained studies on the grammatical of subject and topic in Japanese have had significant impacts on the study of grammatical relations in European as well as Austronesian languages. In the study of word order, the analysis of Japanese numeral quantifiers is used as one of the leading pieces of evidence for the existence of a movement rule in human language. Under case marking, the way subjects are case-marked in Japanese has played a central role in the study of case marking in the Altaic language family. Recent studies of nominalizations have been central to the analysis of their modification and referential functions in a wide variety of languages from around the globe with far-reaching implications to past studies of such phenomena as parts of speech, (numeral) classifiers, and relative clauses. And the study of how in Japanese prosody plays a crucial role in interpretation has become the basis of some important recent developments in the study of wh-questions.

The ***Handbook of Japanese Semantics and Pragmatics*** presents a collection of studies on linguistic meaning in Japanese, either as conventionally encoded in linguistic form (the field of semantics) or as generated by the interaction of form with context (the field of pragmatics). The studies are organized around a model that has long currency in traditional Japanese grammar, whereby the linguistic clause consists of a multiply nested structure centered in a propositional core of objective meaning around which forms are deployed that express progressively more subjective meaning as one moves away from the core toward the periphery of the clause. Following this model, the topics treated in this volume range from aspects of meaning associated with the propositional core, including elements of meaning structured in lexical units (lexical semantics), all the way to aspects of meaning that are highly subjective, being most grounded in the context of the speaker. In between these two poles of the semantics-pragmatics continuum are elements of meaning that are defined at the level of propositions as a whole or between different propositions (propositional logic) and forms that situate propositions in time as events and those situating events in reality including non-actual worlds, e.g., those hoped for (desiderative meaning), denied (negation), hypothesized (conditional meaning), or viewed as ethically or epistemologically possible or necessary (epistemic and deontic modality). Located yet closer to the periphery of the Japanese clause are a rich array of devices for marking propositions according to the degree to which the speaker is committed to their veracity, including means that mark differing perceptual and cognitive modalities and those for distinguishing information variously presupposed.

These studies in Japanese syntax and semantics are augmented by cross-linguistic studies that examine various topics in these fields from the perspectives of language

universals and the comparative study of Japanese and another language. The ***Handbook of Japanese Contrastive Linguistics*** sets as its primary goal uncovering principled similarities and differences between Japanese and other languages around the globe and thereby shedding new light on the universal and language-particular properties of Japanese. Topics ranging from inalienable possession to numeral classifiers, from spatial deixis to motion typology, and from nominalization to subordination, as well as topics closely related to these phenomena are studied in the typological universals framework. Then various aspects of Japanese such as resultative-progressive polysemy, entailment of event realization, internal-state predicates, topic constructions, and interrogative pronouns, are compared and contrasted with individual languages including Ainu, Koryak, Chinese, Korean, Newar, Thai, Burmese, Tagalog, Kapampangan, Lamaholot, Romanian, French, Spanish, German, English, Swahili, Sidaama, and Mayan languages.

6 Psycholinguistics and Applied Linguistics

HJLL includes two volumes containing topics related to wider application of Japanese linguistics and to those endeavors seeking grammar-external evidence for the psycho-neurological reality of the structure and organization of grammar. By incorporating the recent progress in the study of the cognitive processes and brain mechanisms underlying language use, language acquisition, and language disorder, the ***Handbook of Japanese Psycholinguistics*** discusses the mechanisms of language acquisition and language processing. In particular, the volume seeks answers to the question of how Japanese is learned/acquired as a first or second language, and pursues the question of how we comprehend and produce Japanese sentences. The chapters in the acquisition section allow readers to acquaint themselves with issues pertaining to the question of how grammatical features (including pragmatic and discourse features) are acquired and how our brain develops in the language domain, with respect to both language-particular and universal features. Specific topics dealt with include Japanese children's perceptual development, the conceptual and grammatical development of nouns, Japanese specific language impairment, narrative development in the L1 cognitive system, L2 Japanese acquisition and its relation to L1 acquisition. The language processing section focuses on both L1 and L2 Japanese processing and covers topics such as the role of prosodic information in production/comprehension, the processing of complex grammatical structures such as relative clauses, the processing issues related to variable word order, and lexical and sentence processing in L2 by speakers of a different native language.

The ***Handbook of Japanese Applied Linguistics*** complements the Psycholinguistics volume by examining language acquisition from broader sociocultural per-

spectives, i.e., language as a means of communication and social behavioral system, emphasizing pragmatic development as central to both L1 and L2 acquisition and overall language/human development. Topics approached from these perspectives include the role of caregiver's speech in early language development, literacy acquisition, and acquisition of writing skills. Closely related to L1 and L2 acquisition/development are studies of bilingualism/multilingualism and the teaching and learning of foreign languages, including Japanese as a second language, where topics discussed include cross-lingual transfer from L1 to L2, learning errors, and proficiency assessment of second language acquisition. Chapters dealing with topics more squarely falling in the domain of applied linguistics cover the issues in corpus/computational linguistics (including discussions on CHILDES for Japanese and the YK corpus widely-used in research on Japanese as a second language), clinical linguistics (including discussions on language development in children with hearing impairment and other language disorders, with Down syndrome, or autism), and translation and interpretation. Technically speaking, Japanese Sign Language is not a variety of Japanese. However, in view of the importance of this language in Japanese society and because of the rapid progress in sign language research in Japan and abroad and what it has to offer to the general theory of language, chapters dealing with Japanese Sign Language are also included in this volume.

7 Grammatical Sketch of Standard Japanese

The following pages offer a brief overview of Japanese grammar as an aid for a quick grasp of the structure of Japanese that may prove useful in studying individual, thematically organized handbooks of this series. One of the difficult problems in presenting non-European language materials using familiar technical terms derived from the European grammatical tradition concerns mismatches between what the glosses may imply and what grammatical categories they are used to denote in the description. We will try to illustrate this problem below as a way of warning not to take all the glosses at their face value. But first some remarks are in order about the conventions of transcription of Japanese, glossing of examples, and their translations used in this series.

7.1 Writing, alphabetic transcription, and pronunciation

Customarily, Japanese is written by using a mixture of Chinese characters (for content words), *hiragana* (for function words such as particles, suffixes and inflectional endings), *katakana* (for foreign loans and mimetics), and sometimes Roman alphabet.

Because Japanese had no indigenous writing system, it developed two phonogram systems of representing a phonological unit of “mora”, namely *hiragana* and *katakana*, by simplifying or abbreviating (parts of) Chinese characters. *Hiragana* and *katakana* syllabaries are shown in Table 1, together with the alphabetic transcriptions adopted in the HJLL series.

Table 1: *Alphabetic transcriptions adopted in HJLL*

transcription	*a*	*ka*	*sa*	*ta*	*na*	*ha*	*ma*	*ya*	*ra*	*wa*	*n*
hiragana	あ	か	さ	た	な	は	ま	や	ら	わ	ん
katakana	ア	カ	サ	タ	ナ	ハ	マ	ヤ	ラ	ワ	ン
transcription	*i*	*ki*	*si*	*ti*	*ni*	*hi*	*mi*	–	*ri*	–	
hiragana	い	き	し	ち	に	ひ	み	–	り	–	
katakana	イ	キ	シ	チ	ニ	ヒ	ミ	–	リ	–	
transcription	*u*	*ku*	*su*	*tu*	*nu*	*hu*	*mu*	*yu*	*ru*	–	
hiragana	う	く	す	つ	ぬ	ふ	む	ゆ	る	–	
katakana	ウ	ク	ス	ツ	ヌ	フ	ム	ユ	ル	–	
transcription	*e*	*ke*	*se*	*te*	*ne*	*he*	*me*	–	*re*	–	
hiragana	え	け	せ	て	ね	へ	め	–	れ	–	
katakana	エ	ケ	セ	テ	ネ	ヘ	メ	–	レ	–	
transcription	*o*	*ko*	*so*	*to*	*no*	*ho*	*mo*	*yo*	*ro*	*o*	
hiragana	お	こ	そ	と	の	ほ	も	よ	ろ	を	
katakana	オ	コ	ソ	ト	ノ	ホ	モ	ヨ	ロ	ヲ	

Because of phonological change, the columns indicated by strikethroughs have no letters in contemporary Japanese, although they were filled in with special letters in classical Japanese. If all the strikethroughs were filled, the chart will contain 50 letters for each of *hiragana* and *katakana*, so the syllabary chart is traditionally called *Gojū-on zu* (chart of 50 sounds). To these should be added the letter ん or ン representing a moraic nasal [N], on the rightmost column.

The “50-sound chart”, however, does not exhaust the *hiragana* and *katakana* letters actually employed in Japanese, because the basic consonant sounds (*k, s, t, h*) have variants. The sound represented by the letter *h* is historically related to the sound represented by *p*, and these voiceless obstruents (*k, s, t*, and *p*) have their respective voiced counterparts (*g, z, d*, and *b*). Table 2 shows letters for these consonants followed by five vowels.

Table 2: *Letters for voiced obstruents and bilabial [p]*

transcription	*ga*	*za*	*da*	*ba*	*pa*
hiragana	が	ざ	だ	ば	ぱ
katakana	ガ	ザ	ダ	バ	パ
transcription	*gi*	*zi*	*zi*	*bi*	*pi*
hiragana	ぎ	じ	ぢ	び	ぴ
katakana	ギ	ジ	ヂ	ビ	ピ
transcription	*gu*	*zu*	*zu*	*bu*	*pu*
hiragana	ぐ	ず	づ	ぶ	ぷ
katakana	グ	ズ	ヅ	ブ	プ
transcription	*ge*	*ze*	*de*	*be*	*pe*
hiragana	げ	ぜ	で	べ	ぺ
katakana	ゲ	ゼ	デ	ベ	ペ
transcription	*go*	*zo*	*do*	*bo*	*po*
hiragana	ご	ぞ	ど	ぼ	ぽ
katakana	ゴ	ゾ	ド	ボ	ポ

It is important to note that Tables 1 and 2 show the conventional letters and alphabetical transcription adopted by the HJLL series; they are not intended to represent the actual pronunciations of Japanese vowels and consonants. For example, among the vowels, the sound represented as "u" is pronounced as [ɯ] with unrounded lips. Consonants may change articulation according to the following vowels. Romanization of these has been controversial with several competing proposals.

There are two Romanization systems widely used in Japan. One known as the Hepburn system is more widely used in public places throughout Japan such as train stations, street signs, as well as in some textbooks for learners of Japanese. This system is ostensibly easier for foreigners familiar with the English spelling system. The *Kunreishiki* (the cabinet ordinance system) is phonemic in nature and is used by many professional linguists. The essential differences between the two Romanization systems center on palatalized and affricate consonants, as shown in Table 3 below by some representative syllables for which two Romanization renditions differ:

Table 3: *Two systems of Romanization*

Hiragana	IPA	Hepburn	Kunreishiki
し	[ʃi]	shi	si
しゃ	[ʃa]	sha	sya
しゅ	[ʃɯ]	shu	syu
しょ	[ʃo]	sho	syo
じ and ぢ	[ʤi]	ji	zi
じゃ	[ʤa]	ja	zya
じゅ	[ʤɯ]	ju	zyu
じょ	[ʤo]	jo	zyo
ち	[tʃi]	chi	ti
ちゃ	[tʃa]	cha	tya
ちゅ	[tʃɯ]	chu	tyu
ちょ	[tʃo]	cho	tyo
つ	[tsɯ]	tsu	tu
づ and ず	[dzɯ]	dzu	zu
ふ	[ɸɯ]	fu	hu

Except for the volumes on Ryukyuan, Ainu, and Japanese dialects, whose phonetics differ from Standard Japanese, HJLL adopts the Kunreishiki system for rendering cited Japanese words and sentences but uses the Hepburn system for rendering conventional forms such as proper nouns and technical linguistic terms in the text and in the translations of examples.

The cited Japanese sentences in HJLL look as below, where the first line transliterates a Japanese sentence in Kunreishiki Romanization, the second line contains interlinear glosses largely following the Leipzig abbreviation convention, and the third line is a free translation of the example sentence.

(1)	*Taroo*	*wa*	*Ziroo*	*to*	*Tookyoo*	*e*	*it-te*	*kutusita*	*o*	*kat-ta.*
	Taro	TOP	Jiro	COM	Tokyo	ALL	go-GER	sock	ACC	buy-PST

‘Taro went to Tokyo with Jiro and bought socks.’

The orthographic convention of rendering Japanese is to represent a sentence with an uninterrupted sequence of Sino-Japanese characters and *katakana* or *hiragana* syllabaries without a space for word segmentation, as in 太郎は次郎と東京へ行って靴下を買った for (1). In line with the general rules of Romanization adopted in

books and articles dealing with Japanese, however, HJLL transliterates example sentences by separating word units by spaces. The example in (1) thus has 10 words. Moreover, as in *it-te* (go-GERUNDIVE) and *kat-ta* (buy-PAST) in (1), word-internal morphemes are separated by a hyphen whenever necessary, although this practice is not adopted consistently in all of the HJLL volumes. Special attention should be paid to particles like *wa* (topic), *to* 'with' and *e* 'to, toward', which, in the HJLL representation, are separated from the preceding noun or noun phrase by a space (see section 7.3). Remember that case and other kinds of particles, though spaced, form phrasal units with their preceding nouns.

7.2 Word order

As seen in (1), Japanese is a verb-final, dependent-marking agglutinative language. It is basically an SOV language, which marks the nominal dependent arguments by particles (*wa*, *to*, *e*, and *o* above), and whose predicative component consists of a verbal-stem, a variety of suffixes, auxiliary verbs, and semi-independent predicate extenders pertaining to the speech act of predication (see section 7.6). While a verb is rigidly fixed in sentence final position, the order of subject and object arguments may vary depending on pragmatic factors such as emphasis, background information, and cohesion. Thus, sentence (2a) with the unmarked order below, in principle, may vary in multiple ways as shown by some possibilities in (2b)–(2d).

(2)	a.	*Taroo*	*ga*	*Hanako*	*ni*	*Ziroo*	*o*	*syookai-si-ta.*
		Taro	NOM	Hanako	DAT	Jiro	ACC	introducing-do-PST

'Taro introduced Jiro to Hanako.'

b. *Taroo ga* ***Ziroo o*** *Hanako ni syookai-si-ta.*
c. ***Hanako ni*** *Taroo ga Ziroo o syookai-si-ta.*
d. ***Ziroo o*** *Taroo ga Hanako ni syookai-si-ta.*

Adverbs, likewise, can be rather freely placed, though each type of adverbs has its basic position.

(3)	a.	***Saiwainimo***	*Hanako*	*ga*	*gohan*	*o*	*tai-te*	*kure-te*	*i-ta.*
		luckily	Hanako	NOM	rice	ACC	cook-GER	GIVE-GER	BE-PST

'Luckily Hanako had done the favor of cooking the rice (for us).'

b. *Hanako ga* ***saiwainimo*** *gohan o tai-te kure-te i-ta.*
c. *Hanako ga gohan o* ***saiwainimo*** *tai-te kure-te i-ta.*

Notice that while the verbal complex in the sentence above is not as tightly organized as a complex involving suffixes, a sentence adverb cannot be placed within the verbal complex, showing that the sequence of *tai-te kure-te i-ta* forms a tighter constituent,

which, however, permits insertion of the topic particle *wa* after each of the gerundive forms. (See section 7.4 below on the nature of gerundive forms in Japanese.)

As the normal position of sentence adverbs is sentence initial, manner and resultative adverbs have an iconically-motivated position, namely before and after the object noun phrase, respectively, as below, though again these adverbs may move around with varying degrees of naturalness:

(4)	*Hanako*	*ga*	***isoide***	*gohan*	*o*	*tai-te*	*kure-ta.*
	Hanako	NOM	hurriedly	rice	ACC	cook-GER	GIVE-PST

‘Hanako did the favor of cooking the rice hurriedly (for us).’

(5)	*Hanako*	*ga*	*gohan*	*o*	***yawarakaku***	*tai-te*	*kure-ta.*
	Hanako	NOM	rice	ACC	softly	cook-GER	GIVE-PST

‘Hanako did the favor of cooking the rice soft (for us).’

The fact that an object noun phrase can be easily separated from the verb, as in (2b.d), and that adverbs can freely intervene between an object and a verb, as in (5), has raised the question whether Japanese has a verb phrase consisting of a verb and an object noun phrase as a tightly integrated constituent parallel to the VP in English (cf. **cook hurriedly the rice* – the asterisk marks ungrammatical forms).

7.3 NP structure

Noun phrases, when they occur as arguments or adjuncts, are marked by case particles or postpositions that are placed after their host nouns. Because case markers can be set off by a pause, a filler, or even longer parenthetic material, it is clear that they are unlike declensional affixes in inflectional languages like German or Russian. Their exact status, however, is controversial; some researchers regard them as clitics and others as (non-independent) words.

Elaboration of Japanese noun phrases is done by prenominal modifiers such as a demonstrative, a genitive noun phrase, or an adjective, as below, indicating that Japanese is a consistent head-final language at both nominal and clausal levels.

(6)	a.	*kono*	*Taroo*	*no*	*kaban*
		this	Taro	GEN	bag

lit. ‘this Taro’s bag’

	b.	*Taroo*	*no*	*kono*	*kaban*
		Taro	GEN	this	bag

lit. ‘Taro’s this bag’

Japanese lacks determiners of the English type that "close off" NP expansion. The literal translations of the Japanese forms above are ungrammatical, indicating that English determiners like demonstratives and genitive noun phrases do not allow further expansion of an NP structure. Also seen above is the possibility that prenominal modifiers can be reordered just like the dependents at the sentence level. The order of prenominal modifiers, however, is regulated by the iconic principle of placing closer to the head noun those modifiers that have a greater contribution in specifying the nature and type of the referent. Thus, descriptive adjectives tend to be placed closer to a head noun than demonstratives and genitive modifiers of non-descriptive types. Interesting is the pattern of genitive modifiers, some of which are more descriptive and are placed closer to the head noun than others. Genitives of the same semantic type, on the other hand, can be freely reordered. Compare:

(7) a. *Yamada-sensei no kuroi kaban*
Yamada-professor GEN black bag
'Professor Yamada's black bag'
b. **kuroi Yamada-sensei no kaban*
(O.K. with the reading of 'a bag of Professor Yamada who is black')

(8) a. *Yamada-sensei no gengogaku no koogi*
Yamada-professor GEN linguistics GEN lecture
'Professor Yamada's linguistics lecture'
b. **gengogaku no Yamada-sensei no koogi*
(O.K. with the reading of 'a lecture by Professor Yamada of linguistics')

(9) a. *Yamada-sensei no kinoo no koogi*
Yamada-professor GEN yesterday GEN lecture
lit. 'Professor Yamada's yesterday's lecture' 'Yesterday's lecture by Professor Yamada'
b. *Kinoo no Yamada-sensei no koogi*

(10) a. *oomori no sio-azi no raamen*
big.serving GEN salt-tasting GEN *ramen*
lit. 'big-serving salt-tasting *ramen* noodles'
b. *sio-azi no oomori no raamen*

(11) a. *atui sio-azi no raamen*
hot salt-tasting GEN *ramen*
'hot salt-tasting *ramen* noodles'
b. *sio-azi no atui ramen*

Numeral classifiers (CLFs) pattern together with descriptive modifiers so that they tend to occur closer to a head noun than a possessive genitive phrase.

(12) a. *Taroo no san-bon no enpitu*
Taro GEN three-CLF GEN pencil
'Taro's three pencils'
b. **san-bon no Taroo no enpitu*

Numeral classifiers also head an NP, where they play a referential function and where they can be modified by a genitive phrase or an appositive modifier, as in (13a.b). They may also "float" away from the head noun and become adverbial, as in (13c).

(13) a. *Taroo wa gakusei no* ***san-nin*** *o mikake-ta.*
Taro TOP student GEN three-CLF ACC see.by.chance-PST
'Taro saw three of students by chance.'

b. *Taroo wa gakusei* ***san-nin*** *o mikake-ta.*
Taro TOP student three-CLF ACC see.by.chance-PST
lit. 'Taro saw student-threes by chance.'

c. *Taroo wa gakusei o* ***san-nin*** *mikake-ta.*
Taro TOP student ACC three-CLF see.by.chance-PST
'Taro saw students, three (of them), by chance.'

As in many other SOV languages, the so-called relative clauses are also prenominal and are directly placed before their head nouns without the mediation of "relative pronouns" like the English *which* or *who* or "complementizers" like *that*. The predicates in relative clauses are finite, taking a variety of tense and aspect. The subject may be replaced by a genitive modifier. Observe (14a).

(14) a. *Boku mo [Taroo ga/no kat-ta] hon o kat-ta.*
I ADVPART Taro NOM/GEN buy-PST book ACC buy-PST
'I also bought the book which Taro bought.'

b. *Boku mo [Taroo ga/no kat-ta] no o kat-ta.*
I ADVPART Taro NOM/GEN buy-PST NM ACC buy-PST
'I also bought the one which Taro bought.'

The structure used as a modifier in the relative clause construction can also head a noun phrase, where it has a referential function denoting an entity concept evoked by the structure. In Standard Japanese such a structure is marked by the nominalization particle *no*, as in (14b).

7.4 Subject and topic

Some of the sentences above have noun phrases marked by the nominative case particle *ga* and some by the topic marker *wa* for what appear to correspond to the subject noun phrases in the English translations. This possibility of *ga*- and *wa*-marking is seen below.

(15) a. *Yuki* *ga* *siro-i.*
snow NOM white-PRS
'The snow is white.'

b. *Yuki* *wa* *siro-i.*
snow TOP white-PRS
'Snow is white.'

As the difference in the English translations indicates, these two sentences are different in meaning. Describing the differences between topic and non-topic sentences has been a major challenge for Japanese grammarians and teachers of Japanese alike. The difference in the English translations above, however, is indicative of how these two sentences might differ in meaning. Sentence (15a) describes a state of affairs involving specific snow just witnessed, whereas (15b) is a generic statement about a property of snow unbounded by time. Thus, while (15a) would be uttered only when the witnessed snow is indeed white, (15b) would be construed true even though we know that there are snow piles that are quite dirty.

A similar difference is seen in verbal sentences as well.

(16) a. *Tori* *ga* *tob-u.*
bird NOM fly-PRS
'A bird is flying/is about to fly.'

b. *Tori* *wa* *tob-u.*
bird TOP fly-PRS
'Birds fly.'

Non-topic sentences like (15a) and (16a) are often uttered with an exclamation accompanying a sudden discovery of a state of affairs unfolding right in front of one's eyes. The present tense forms (*-i* for adjectives and *-(r)u* for verbs) here anchor the time of this discovery to the speech time. The present tense forms in (15b) and (16b), on the other hand, mark a generic tense associated with a universal statement.

These explanations can perhaps be extended to a time-bound topic sentence seen in (17b) below.

(17) a. *Taroo ga hasit-ta.*
Taro NOM run-PST
'Taro NOM ran.'

b. *Taroo wa hasit-ta.*
Taro TOP run-PST
'Taro ran.'

That is, while (17a) reports an occurrence of a particular event at a time prior to the speech time, (17b) describes the nature of the topic referent – that Taro was engaged in the running activity – as a universal truth of the referent, but universal only with respect to a specifically bound time marked by the past tense suffix.

Topics need not be a subject, and indeed any major sentence constituent, including adverbs, may be marked topic in Japanese, as shown below.

(18) a. *Sono hon wa Taroo ga yon-de i-ru.*
that book TOP Taro NOM read-GER BE-PRS
'As for that book, Taro is reading (it).'

b. *Kyoo wa tenki ga yo-i.*
today TOP weather NOM good-PRS
'As for today, the weather is good.'

c. *Sonnani wa hayaku wa hasir-e na-i.*
that.way TOP quickly TOP run-POTEN NEG-PRS
'That quickly, (I) cannot run.'

7.4 Complex sentences

As in many Altaic languages, compound sentences in Japanese do not involve a coordinate conjunction like English *and*. Instead, clauses are connected by the use of inflected verb forms, as in (19a) below, where the *-i* ending is glossed in the HJLL series as either INF (infinitive) or ADVL (adverbal) following the Japanese term *ren'yō-kei* for the form. While the *-i* ending in the formation of compound sentences is still used today, especially in writing, the more commonly used contemporary form involves a conjunctive particle *-te* following the *-i* infinitive form, as in (19b) below. In HJLL, this combination is glossed as GER (gerundive), though the relevant Japanese forms do not have the major nominal use of English gerundive forms.

(19) a. *Hana wa sak-i, tori wa uta-u.*
flower TOP bloom-INF bird TOP sing-PRS
'Flowers bloom and birds sing.'

b. *Hana wa sa.i-te, tori wa uta-u.*
flower TOP bloom-GER bird TOP sing-PRS
'Flowers bloom and birds sing.'

Both the *-i* and *-te* forms play important roles in Japanese grammar. They are also used in clause-chaining constructions for serial events (20a), and in complex sentences (20b)–(20d), as well as in numerous compound verbs (and also in many compound nouns) such as *sak-i hokoru* (bloom-INF boast) 'be in full bloom', *sak-i tuzukeru* (bloom-INF continue) 'continue blooming', *sa.i-te iru* (bloom-GER BE) 'is blooming', and *sa.i-te kureru* (bloom-GER GIVE) 'do the favor of blooming (for me/us)'.

(20) a. *Taroo wa [ok-i/ok.i-te], [kao o ara-i/arat-te],*
Taro TOP rise-INF/rise-GER face ACC wash-INF/wash-GER
[gohan o tabe-ta].
meal ACC eat.PST
'Taro got up, washed his face, and ate a meal.'

b. *Taroo wa [sakana o tur-i] ni it-ta.*
Taro TOP fish ACC catch-INF DAT go-PST
'Taro went to catch fish.'

c. *Taroo wa [aruk-i nagara] hon o yon-da.*
Taro TOP walk-INF SIMUL book ACC read-PST
'Taro read a book while walking.'

d. *Taroo wa [Hanako ga ki-ta no] ni awa-na-katta.*
Taro TOP Hanako NOM come-PST NM DAT see-NEG-PST.
'Taro did not see (her), even though Hanako came.'

(20d) has the nominalized clause marked by the particle *no* followed by the dative *ni*, also seen in (20b) marking the purposive form. Now the *no-ni* sequence has been reanalyzed as a concessive conjunction meaning 'even though'.

7.5 Context dependency

The context dependency of sentence structure in Japanese is much more clearly pronounced than in languages like English. Indeed, it is rare that Japanese sentences express all the arguments of a verb such as a subject (or topic) and an object noun phrase included in the sentences used above for illustrative purposes. A typical dialog would take the following form, where what is inferable from the speech context is not expressed.

(21) a. Speaker A: *Tokorode, Murakami Haruki no saisin-saku yon-da ka.*
by.the.way Murakami Haruki GEN newest-work read-PST Q
'By the way, have (you) read Haruki Murakami's latest work?'

b. Speaker B: *Un, moo yon-da.*
uh-hu already read-PST
'Uh-hu, (I) already read (it)'.

In (21a) A's utterance is missing a subject noun phrase referring to the addressee, and B's response in (21b) is missing both subject and object noun phrases. In some frameworks, sentences like these are analyzed as containing zero pronouns or as involving a process of "pro drop", which deletes assumed underlying pronouns. This kind of analysis, however, ignores the role of speech context completely and incorporates information contextually available into sentence structure. In an analysis that takes seriously the dialogic relationship between speech context and sentence structure, the expressions in (21) would be considered full sentences as they are.

7.6 Predicative verbal complexes and extenders

Coding or repeating contextually determinable verb phrases, as in (21b), is less offensive than expressing contextually inferable noun phrases presumably because verb phrases have the predication function of assertion, and because they also code a wide range of other types of speech acts and of contextual information pertaining to the predication act. Declarative sentences with plain verbal endings like the one in (21b) are usable as "neutral" expressions in newspaper articles and literary works, where no specific reader is intended. In daily discourse, the plain verbal forms "explicitly" code the speaker's attitude toward the hearer; namely, that the speaker is treating the hearer as his equal or inferior in social standing, determined primarily by age, power, and familiarity. If the addressee were socially superior or if the occasion demanded formality, a polite, addressee honorific form with the suffix *-masu* would be used, as below.

(22) *Hai, moo yom-i-masi-ta.*
yes already read-INF-POL-PST
'Yes, (I have) already read (it).'

The referent honorific forms are used when the speaker wishes to show deference toward the referent of arguments – subject honorific and object honorific (or humbling) forms depending on the type of argument targeted. If (21b) were to be uttered in reference to a social superior, the following would be more appropriate:

(23) *Un, (Yamada-sensei wa) moo yom-are-ta.*
uh-hu (Yamada-professor TOP) already read-SUB.HON-PST
'Uh-hu, (Professor Yamada has) already read (it).'

This can be combined with the polite ending *-masu*, as below, where the speaker's deference is shown to both the referent of the subject noun phrase and the addressee:

(24) *Hai, (Yamada-sensei wa) moo yom-are-masi-ta.*
Yes (Yamada-professor TOP) already read-HON-POL-PST
'Yes, (Professor Yamada has) already read (it).'

As these examples show, Japanese typically employs agglutinative suffixes in the elaboration of verbal meanings associated with a predication act. The equivalents of English auxiliary verbs are either suffixes or formatives connected to verb stems and suffixed forms in varying degrees of tightness. These are hierarchically structured in a manner that expresses progressively more subjective and interpersonal meaning as one moves away from the verb-stem core toward the periphery. For example, in the following sentence a hyphen marks suffixal elements tightly bonded to the preceding form, an equal sign marks a more loosely connected formative, which permits insertion of certain elements such as the topic particle *wa*, and a space sets off those elements that are independent words following a finite predicate form, which may terminate the utterance.

(25) *(Taroo wa) ik-ase-rare-taku=na-katta rasi-i mitai des-u wa.*
(Taro TOP) go-CAUS-PASS-DESI=NEG-PST CONJEC-PRS UNCERT POLCOP-PRS SFP
'(Taro) appears to seem to not want to have been forced to go, I tell you.'

The final particle *wa* above encodes the information that the speaker is female. A male speaker would use *yo* or *da yo*, the latter a combination of the plain copula and *yo*, instead of *desu wa* above, or combinations such as *da ze* and *da zo* in rough speech.

Non-declarative Japanese sentences, on the other hand, frequently suppress auxiliary verbs, the copula, and the question particle especially in casual speech, where intonation and tone of voice provide clues in guessing the intended speech act. Casual interrogatives take the form of (26a) with a nominalization marker bearing a rising intonation, marked by the question mark in the transcription, whereas fuller versions have the interrogative particle *ka* or a combination of the polite copula and *ka*, as in (26b).

(26) a. *Moo kaeru no?*
already return NM
'Going home already?'

b. *Moo* *kaeru* *no* *(desu)* *ka.*
already return NM (POLCOP) Q
'Going home already?'

Requests are made with the aid of an auxiliary-like "supporting" verb *kureru* 'GIVE (ME THE FAVOR OF. . .)', its polite form *kudasai*, or its intimate version *tyoodai*, as seen in (27a). Again, these forms are often suppressed in a highly intimate conversation and may result in a form like (27b).

(27) a. *Hayaku* *kaet-te* *kure/kudasai/tyoodai.*
soon return-GER GIVE/GIVE.POL/GIVE.INTI
'(Please) come home soon (for me/us).'

b. *Hayaku* *kaet-te* *ne.*
soon return-GER SFP
'(Please) come home soon, won't you?'

The use of dependent forms (e.g., the gerundive *-te* form above) as independent sentences is similar to that of subjunctive forms of European languages as independent sentences, as illustrated by the English sentence below.

(28) *If you would give me five thirty-cent stamps.*

Conditionals are used as independent suggestion sentences in Japanese as well. For example, (29a) has a fuller version like (29b) with the copula as a main-clause verb, which can also be suppressed giving rise to the truncated form (29c).

(29) a. *Hayaku* *kaet-tara?*
quickly return-COND
lit. ' If return quickly.' 'Why don't you go home quickly?'

b. *Hayaku* *kaet-tara* *ikaga* *desu* *ka.*
quickly return-COND how POLCOP Q
lit. 'How is it if (you) went home quickly?'

c. *Hayaku* *kaet-tara* *ikaga?*
quickly return-COND how
'Why don't (you) go home quickly?'

Understanding Japanese utterances requires full recourse to the elements of speech context, such as the nature of the speaker and the hearer and the social relationship between them, the information "in the air" that is readily accessible to the interlocutors, and the formality of the occasion. Indeed, the difficult part of the art of

speaking Japanese is knowing how much to leave out from the utterance and how to infer what is left unsaid.

8 Conclusion

Many of the interesting topics in Japanese grammar introduced above are discussed in great detail in the Lexicon-Word formation handbook and the Syntax volume. The Historical handbook also traces developments of some of the forms and constructions introduced above. The Sociolinguistics volume gives fuller accounts of the sentence variations motivated by context and discourse genre.

References

Chamberlain, Basil H. 1895. *Essay in aid of a grammar and dictionary of the Luchuan language*. Transactions of the Asiatic Society of Japanese, vol. 23 supplement.
Frellesvig, Bjarke. 2010. *A history of the Japanese language*. Cambridge: Cambridge University Press.
Martin, Samuel E. 1975. *A reference grammar of Japanese*. New Haven: Yale University Press.
Miller, Roy A. 1971. *Japanese and the other Altaic languages*. Chicago: University of Chicago Press.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Appendix: List of abbreviations for HJLL

1	first person
2	second person
3	third person
A	agent-like argument of canonical transitive verb
ABL	ablative
ACC	accusative
ACOP	adjectival copula
ADJ	adjective
AND	adnominal
ADV	adverb(ial(izer))
ADVL	adverbal
ADVPART	adverbial particle
AGR	agreement
AGT	agent
ALL	allative
AN	adjectival noun

ANTIP	antipassive
AP	adverbial particle, adjective phrase
APPL	applicative
ART	article
ASP	aspect
ATTR	attributive
AUX	auxiliary
AUXV	auxiliary verb
C	consonant
CAUS	causative
CLF	classifier
COHORT	cohortative
COM	comitative
COMP	complementizer
COMPL	completive
CONC	concessive
CONCL	conclusive
COND	conditional
CONJEC	conjectural
CONJCT	conjunctive
CONT	continuative
COP	copula
CVB	converb
DAT	dative
D	demonstrative
DECL	declarative
DEF	definite
DEM	demonstrative
DET	determiner
DESI	desiderative
DIST	distal
DISTR	distributive
DO	direct object
DU	dual
DUR	durative
EMPH	emphatic
ERG	ergative
ETOP	emphatic topic
EVID	evidential
EXCL	exclamatory, exclusive
EXPL	expletive
FOC	focus

FUT	future
GEN	genitive
GER	gerund(ive)
H	high (tone or pitch)
HON	honorific
HUM	humble
IMP	imperative
INCL	inclusive
IND	indicative
INDEF	indefinite
INF	infinitive
INS	instrumental
INT	intentional
INTERJEC	interjection
INTI	intimate
INTR	intransitive
IO	indirect object
IRR	irrealis
ITERA	iterative
k-irr	k-irregular (*ka-hen*)
L	low (tone or pitch)
LB	lower bigrade (*shimo nidan*)
LM	lower monograde (*shimo ichidan*)
LOC	locative
MPST	modal past
MVR	mid vowel raising
N	noun
n-irr	n-irregular (*na-hen*)
NCONJ	negative conjectual
NEC	neccessitive
NEG	negative
NM	nominalization marker
NMLZ	nominalization/nominalizer
NMNL	nominal
NOM	nominative
NONPST	nonpast
NP	noun phrase
OBJ	object
OBL	oblique
OPT	optative
P	patient-like argument of canonical transitive verb, preposition, postposition

PART	particle
PASS	passive
PCONJ	present conjectural
PERF	perfective
PL	plural
POL	polite
POLCOP	polite copula
POSS	possessive
POTEN	potential
PP	prepositional/postpositional phrase
PRED	predicative
PRF	perfect
PRS	present
PRES	presumptive
PROG	progressive
PROH	prohibitive
PROV	provisional
PROX	proximal/proximate
PST	past
PSTCONJ	past conjectural
PTCP	participle
PURP	purposive
Q	question/question particle/question marker
QD	quadrigrade (*yodan*)
QUOT	quotative
r-irr	r-irregular (*ra-hen*)
REAL	realis
RECP	reciprocal
REFL	reflexive
RES	resultative
RESP	respect
S	single argument of canonical intransitive verb, sentence
SBJ	subject
SBJV	subjunctive
SFP	sentence final particle
SG	singular
SIMUL	simultaneous
s-irr	s-irregular (*sa-hen*)
SG	singular
SPON	spontaneous
SPST	simple past
STAT	stative

TOP	topic
TR	transitive
UB	upper bigrade (*kami-nidan*)
UNCERT	uncertain
UM	upper monograde (*kami-ichidan*)
V	verb, vowel
VN	verbal noun
VOC	vocative
VOL	volitional
VP	verb phrase

LANGUAGES

ConJ	contemporary Japanese
EMC	Early Middle Chinese
EMJ	Early Middle Japanese
EOJ	Eastern Old Japanese
J-Ch	Japano-Chinese
LMC	Late Middle Chinese
LMJ	Late Middle Japanese
JPN	Japanese
MC	Middle Chinese
MJ	Middle Japanese
MK	Middle Korean
ModJ	Modern Japanese
OC	Old Chinese
OJ	Old Japanese
pJ	proto-Japanese
pK	proto-Korean
SJ	Sino-Japanese
Skt	Sanskrit

Table of contents

Contributors

Masako FUJIMOTO is a Research Fellow at the National Institute for Japanese Language and Linguistics (NINJAL). She earned her PhD in Medicine from the University of Tokyo with her work at the Research Institute of Logopedics and Phoniatrics. Her research interests mainly focus on physiological and acoustic characteristics of Japanese phonemes, especially on devoiced vowels and *sokuon*, with reference to those of other languages. She also works on paralinguistic aspects of speech and Southern Ryukyuan dialects. Her recent papers have appeared in *Acoustical Science and Technology* (2013) and *Sophia Linguistica* 60 (2013).

Yukari HIRATA is Associate Professor of Japanese and Linguistics in the Department of East Asian Languages and Literatures at Colgate University. Her research in phonetics and second language speech acquisition investigates how factors such as speaking rate and multi-modal input affect native and non-native speakers' perception and production of Japanese length contrasts. Her recent publications include 'Effects of lips and hands on auditory learning of second-language speech sounds' (with Spencer D. Kelly, *Journal of Speech, Language, and Hearing Research*, 2010).

Yosuke IGARASHI is Associate Professor of Linguistics at Hiroshima University. He received his PhD in linguistics from the Tokyo University of Foreign Studies in 2005. He has worked at the National Institute for Japanese Language, the Laboratory for Language Development, RIKEN Brain Science Institute, and the Japan Society for the Promotion of Science. His research interests cover a range of topics in prosody of languages in general. Recent publications include 'Typology of intonational phrasing in Japanese dialects' in *Prosodic Typology II*, Oxford University Press, 2014.

Shinichiro ISHIHARA is a Research Associate at Johann Wolfgang Goethe University Frankfurt am Main. He has investigated various phenomena at the syntax-prosody interface and their relation to information structure in Japanese and other languages. Recent publications include 'Japanese focus prosody revisited: Freeing focus from prosodic phrasing' (*Lingua* 121, 2011) and 'Syntactic metamorphosis: Clefts, sluicing, and in-situ focus in Japanese' (with Ken Hiraiwa, *Syntax* 15, 2012). He is currently editing (with Caroline Féry) *Oxford Handbook of Information Structure*, to be published by Oxford University Press.

Junko ITO is Professor of Linguistics at UC Santa Cruz. Since her dissertation and related work ('Prosodic theory of epenthesis', *NLLT*, 1989), her research has been concerned with constraint-based phonological theory. In recent years, she has pursued this work within an optimality-theoretic model of phonology ('Recursive prosodic phrasing in Japanese' in *Prosody Matters*, Equinox, 2012). A secondary line of investigation concerns the structure of the phonological lexicon and its implications for the theory of grammar ('Lexical classes in phonology', in *The Oxford Handbook of Japanese Linguistics*. OUP, 2008).

Itsue KAWAGOE is Professor of Linguistics in the Department of Foreign Studies at Kyoto Sangyo University. Her research interests focus on the comparative study of English and Japanese phonetics and phonology, especially on the adaptation of loanwords from English into Japanese. Her recent publications include 'Geminate judgments of English-like words by Japanese native speakers: Differences in the borrowed forms of "stuff" and "tough"' (with Akiko Takemura, *Journal of East Asian Linguistics* 22, 2013) and 'Onsetsu to mōra' [Syllables and moras] in *On'in'ron* [*Phonology*], Asakura Publishsers, 2014.

Shigeto KAWAHARA completed his PhD at the University of Massachusetts, Amherst in 2007. Beginning in 2008 he taught at Rutgers University, where he was a director of the phonetics laboratory. He is now Senior Assistant Professor at the Keio Institute of Linguistic and Cultural Studies. He has worked on both phonetics and phonology, and their interface, most extensively on Japanese, but also as it relates to other languages. He has published in many journals including *Language*, *NLLT, JIPA, Journal of Phonetics, Phonetica*, *Attention*, *Perception & Psychophysics*, and *Journal of Child Language*.

Haruo KUBOZONO completed his PhD at the University of Edinburgh in 1988. He taught phonetics and phonology at Nanzan University, Osaka University of Foreign Studies and Kobe University before he moved to the National Institute for Japanese Languages and Linguistics as Professor/Director in 2010. His research interests range from speech disfluencies to speech prosody (accent and intonation) and its interfaces with syntax and information structure. He recently edited special issues on pitch accent and geminate consonants in *Lingua* (2012) and *Journal of East Asian Linguistics* (2013), respectively.

Kikuo MAEKAWA is Professor/Director in the Department of Corpus Studies at NINJAL. His main interest is in the phonetic and linguistic analyses of spontaneous Japanese. For this goal, he has constructed, with his colleagues, large-scale corpora including the *Corpus of Spontaneous Japanese* (2004) and *Balanced Corpus of Contemporary Written Japanese* (2011). His recent publications cover the topics of corpus design (*Lang. Resources and Evaluation*, 2014), compilation of a Japanese frequency dictionary (Routledge 2013, with Y. Tone and M. Yamazaki), and corpus-based analyses of phonetic variations in spontaneous Japanese.

Armin MESTER, Professor of Linguistics at UC Santa Cruz, works on the principles that underlie the prosodic organization of human language, as manifested in systems of syllabification, stress, and accent, in the canonical prosodic forms in word formation, and in syntax-prosody mapping. He is pursuing this work within Optimality Theory, with an additional interest in the basic architecture of this theory. His work includes studies of Classical Latin, German and Japanese: 'The quantitative trochee in Latin' (*NLLT*, 1994), *Japanese Morphophonemics: Markedness and Word Structure* (MIT Press, 2003), and 'Prosodic subcategories in Japanese' (*Lingua*, 2013).

Akio NASU is an Associate Professor at the University of Tsukuba. He has been investigating phonological properties of mimetic words in Japanese. His recent research interests also extend to accentual variations observed in Tokyo Japanese spoken by the younger generation. His publications include 'Phonological markedness and asymmetries in Japanese mimetics' in Haruo Kubozono (ed.), *Asymmetries in Phonology: An East-Asian Perspective* (Kurosio, 2008) and 'Onomatope no gomatsu sokuon' [The word-final moraic obstruent in Japanese mimetics], *Journal of the Phonetic Society of Japan* 11(1), 2007.

Mitsuhiko OTA is a Reader in the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh. His research focuses on phonological development in first and second language as well as in atypical populations. Recent articles include 'Input frequency and lexical variability in phonological development: A survival analysis of word-initial cluster production' (with Sam J. Green, *Journal of Child Language*, 2013,) and 'Revisiting the phonological deficit in dyslexia: Are implicit non-orthographic representations impaired?' (with Catherine Dickie and Ann Clark, *Applied Psycholinguistics*, 2013).

Takashi OTAKE is Director of the E-Listening Laboratory. He completed his PhD at the University of Texas, Austin in 1987. His research interests include issues in phonetics, phonology and spoken-word recognition in Japanese. He has published journal articles in *Journal of Memory and Language*, *Journal of Phonetics*, *Journal of the Acoustical Society of America* and *Language and Speech* (with Anne Cutler). He edited with Anne Cutler *Phonological Structure and Language Processing: Cross-linguistic Studies* (Mouton de Gruyter, 1996).

Gábor PINTÉR is an Associate Professor in the School of Languages and Communication at Kobe University. His research interests lie at the intersection of phonological theory, speech perception and language change. He is currently investigating the role of prosodic information in second language acquisition. His recent publications include 'Boundary and prominence perception by Japanese learners of English – A preliminary study' (*Phonological Studies* 17, 2014). Beyond theoretical topics, he has a long-standing interest in the question of how phonological knowledge can efficiently be utilized in automatic speech recognition systems.

Tomoaki TAKAYAMA is Professor of Japanese Linguistics in the Institute of Human and Social Sciences at Kanazawa University. His research interests focus on historical studies of Japanese phonology, especially phonotactic changes and consonantal issues including *rendaku*. He is the author of *Nihongo on'inshi no dōteki shosō to Kenshukuryōkoshū* [Dynamic aspects of phonological history of Japanese and 'Kenshukuryōkoshū'] (Kasama Shoin, 2014), in which he discusses affrication, consonantal merger, and de-prenasalisation in late Middle Japanese.

Timothy J. VANCE is a Professor in the Department of Linguistic Theory and Structure at the National Institute for Japanese Language and Linguistics in Tokyo, where he is currently heading a large-scale collaborative project on *rendaku*. His research interests include phonology/phonetics, with a focus on Japanese, and writing systems. He is the author of *An Introduction to Japanese Phonology* (SUNY Press, 1987) and *The Sounds of Japanese* (Cambridge University Press, 2008).

Haruo Kubozono

I Introduction to Japanese phonetics and phonology

1 Goals and scope of the volume

This volume describes the basic phonetic and phonological structures of modern Japanese with main focus on the standard variety known as Tokyo Japanese. It aims to provide a comprehensive overview and descriptive generalizations of major phonetic and phonological phenomena in modern Tokyo Japanese by reviewing important studies in the fields over the past century or so. In addition, this volume also aims to give an overview of major phonological theories including, but not restricted to, traditional generative phonology, lexical phonology, prosodic morphology, intonational phonology, and the more recent Optimality Theory. It also presents summaries of interesting questions that remain unsolved in the literature.

While the entire volume is devoted to the description of modern Tokyo Japanese, some chapters refer to other major dialects and some discuss the historical aspects of the dialect to some extent. These references are intended to help the reader better understand the phonetic/phonological structures of modern Tokyo Japanese. It is recommended that one reads the Dialect and History volumes, too, if one is interested in dialects other than Tokyo Japanese and in the history of the language.

This volume consists of eighteen chapters in addition to this introductory chapter to the whole volume (Part I). The eighteen chapters are grouped into four parts from Part II to Part V, according to the nature of the phenomena they deal with. Part II consists of five chapters all of which analyze segmental properties of Japanese such as *sokuon* (or geminate obstruents), new consonant phonemes, vowel devoicing and diphthongs. Part III discusses morphophonological processes and phonetic/phonological structures therein. These include various processes in mimetic, Sino-Japanese, and loanword phonology, various word-formation processes, and *rendaku* (or sequential voicing). Part IV deals with the prosodic structure of modern Japanese, discussing phonetic and phonological processes within and beyond the word. These include word accent, rhythm, and intonation as well as the syntax-phonology interface. Finally, Part V examines Japanese phonetics and phonology from broader perspectives in their interface with other subfields of linguistics such as historical and corpus linguistics, L1 phonology, and L2 research. All four chapters in this part attempt to reveal phonological structures of modern Japanese which would otherwise remain uncovered.

While this volume is intended to cover all major areas in the phonetics and phonology of modern Tokyo Japanese, the reader is referred to more classic or introductory books of Japanese phonetics and phonology including, but not restricted to,

Bloch (1950), Martin (1975), Kawakami (1977), Komatsu (1981), Vance (1987, 2008), Saito (1997), Kubozono (1999b) and Labrune (2012). These introductory books provide basic information and will help the reader to understand the chapters of this volume better.

In this introductory chapter, we will provide an overall introduction to the sound system of modern Tokyo Japanese, such as vowel and consonant inventories, accentual patterns, as well as a brief introduction to each topic dealt with in the subsequent chapters. Also some basic concepts/notions and terminologies that are commonly used in the volume are defined. These include notions such as the mora, the syllable, word accent, and intonation. In addition are sketched the basic organization of the lexicon of modern Japanese and its lexical categories (i.e. native, Sino-Japanese and foreign words).

2 Vowel inventory

2.1 Short vowels

As is well known, modern Tokyo Japanese has five vowel phonemes: /a/, /i/, /u/, /e/ and /o/. In this volume, these vowels are phonetically represented as [a], [i], [ɯ], [e], and [o], respectively. The symbol [ɯ] is used instead of [u] since this vowel has almost lost lip protrusion although it is not as flat as what the IPA symbol [ɯ] is supposed to denote. Phonologically, these five vowels can be characterized by the height of the tongue – high (/i/ and /u/), mid (/e/ and /o/) and low (/a/) – and the backness of the tongue – front (/i/ and /e/) and back (/u/, /o/ and /a/). In terms of distinctive features, the following three features suffice to differentiate between the five vowels (Kubozono 1999b). Positing these features is instrumental in formulating vowel changes such as vowel coalescence (see Kubozono Ch. 5, this volume) and in defining the patterns of vowel epenthesis in loanwords (Kubozono Ch. 8, this volume).

(1) /i/ [+high, -low, -back]
    /u/ [+high, -low, +back]
    /e/ [-high, -low, -back]
    /o/ [-high, -low, +back]
    /a/ [-high, +low, +back]

According to UPSID (UCLA Phonological Segment Inventory Database, Maddieson 1984), a vowel system with five short vowels represents the most standard type of vowel system in world's languages: it is more common than vowel systems with four short vowels or those with six short vowels. The same database also shows

that most five-vowel systems consist of /a/, /i/, /u/, /e/ and /o/. In this sense, the vowel system of modern Japanese represents the most typical vowel system.

The five vowels constituting the vowel system do not occur equally frequently in the Japanese vocabulary. According to Onishi (1937) cited in Hayashi (1982), /a/ is the vowel that occurs most frequently (by type frequency), followed by /o/, /i/, /e/, /u/ in this order.

Of the five vowels, /i/ and /u/ are phonetically 'weak' vowels. They are, for example, the shortest vowels in modern Japanese (Campbell 1992) and are most prone to vowel devoicing (see Fujimoto, this volume, for details). Thus, the second vowels in /a.ki/ 'autumn' and /a.ku/ 'to open (intransitive)' are often devoiced in spontaneous speech, while the second vowels in /a.ka/ 'red', /ta.ke/ 'bamboo' and /ta.ko/ 'octopus, kite' are not. Moreover, /i/ and /u/ are the most common epenthetic vowels in the language. In old loanwords borrowed from Chinese, i.e. Sino-Japanese (SJ) vocabulary, these vowels were inserted to avoid closed syllables, or syllables that end in a consonant. Some examples are given in (1), where old Chinese forms are given in the input and loanword forms in the output. Epenthetic vowels are enclosed in < >.

(2) a. gak → gak<u> 'learning' (学)
yak → yak<u> 'duty' (役)
bat → bat<u> 'punishment' (罰)

b. ek → ek<i> 'benefit' (益), 'station' (駅)
bat → bat<i> 'punishment' (罰)

/i/ and /u/ are inserted in modern loanwords from English and other languages, too. This is exemplified in (2): see Ito and Mester (Ch. 9, this volume) for more details about epenthetic vowels, and Kawahara (Ch. 1, this volume) and Kawagoe (this volume) for the discussion of geminate consonants in modern loanwords.

(3) a. back → bak.k<u>
top → top.p<u>

b. ink → ink<i>, ink<u>
deck → dek.k<i>

2.2 Long vowels and diphthongs

In addition to the five short vowels, modern Tokyo Japanese has the same number of corresponding long vowels. In this volume, these long vowels are represented with a double letter in phonemic/phonological representations (/aa/, /ii/, /uu/, /ee/, and /oo/) and with a length marker in phonetic representations ([a:], [i:], [ɯ:], [e:], [o:]). Not surprisingly, there are many minimal pairs of words that contrast in vowel

length (some of which can be distinguished by word accent). Syllable boundaries are denoted by dots (.) in (4) and in the rest of this volume.

(4) a. too 'ten, tower' vs. to 'door'
b bii.ru 'beer' vs. bi.ru 'building'
c. ru.bii 'ruby (jewel)' vs. ru.bi 'ruby, an old printing type size equal to 5 1/2 points'
d. o.baa.san 'grandmother, old woman' vs. o.ba.san 'aunt, middle-aged woman'
e. paa.maa 'Palmer (personal name)' vs. paa.ma 'permanent wave, perm'
f. koo.mo.ri 'bat' vs. ko.mo.ri 'baby-sitting'
g. oo.ya.ma 'Ōyama (family name)' vs. o.ya.ma 'Oyama (family name)'

Just as their short counterparts, long vowels do not occur equally frequently in the language but variably occur according to the type of the word. First, the main sources of long vowels in modern Japanese are loanwords from Chinese, i.e. SJ words, and recent loanwords from English and other Western languages (henceforth, 'loanwords' for short). That is, long vowels are generally rare in native words. Historically, old Japanese did not have a contrast in vowel length, and developed long vowels via sound changes, notably consonant deletion and vowel coalescence.

(5) a. ka.wa + mo.ri → koo.mo.ri 'bat' (川守り, 蝙蝠)
a.wa + u.mi → oo.mi 'Oomi, Shiga Prefecture' (近江)
b. o.wo.ki.ki → oo.kii 'big'
ko.wi → koi 'carp'

In contrast, there are many SJ morphemes with a long vowel. However, they exhibit a considerable imbalance as regards their distribution: they permit only three long vowels – /oo/, /ee/ and /uu/ – and not /aa/ or /ii/. (6) gives some compounds consisting of two Chinese morphemes/letters.

(6) a. koo.koo 'high school' (高校)
roo.doo 'labor' (労働)
b. see.tee 'establishment' (制定)
tee.see 'correction' (訂正)
c. kuu.ki 'air' (空気)
guu.suu 'even numbers' (偶数)

Of these three vowels, /oo/ is by far the most common in SJ morphemes, followed by /ee/ and /uu/ in this order. Both the absence of /aa/ and /ii/ and the unequal distribution of /oo/, /ee/ and /uu/ in the SJ vocabulary can be attributed primarily to

their historical origins, namely, the fact that long vowels in this type of vocabulary derive largely from diphthongs via vowel coalescence. Thus, /oo/ results from the coalescence of three diphthongs, /au/, /ou/ and /eu/, whereas /ee/ developed from /ai/ and /ei/. The coalescence rule did not yield /aa/ or /ii/ from any vowel sequence (see Kubozono Ch. 5, this volume, for details).

While SJ thus displays a systemic gap in the inventory of long vowels, recent loanwords show all five long vowels because, being sensitive to vowel length, modern Japanese adopts long or tense vowels in the source languages as long vowels.[1] Some examples are given below.

(7) kaa.do 'card'
rii.daa 'leader, reader'
puu.ru 'pool'
tee.bu.ru 'table'
koo.naa 'corner'

It is probably worth mentioning the phonetic differences between long and short vowels. As their names suggest, these two categories are differentiated from each other in terms of phonetic duration: other things being equal, long vowels are considerably longer than their short counterparts. Han (1962: 65) reports that long vowels are phonetically two to three times as long as corresponding short vowels in the same phonological contexts. Hirata (2004) provides more recent data for both accented and unaccented words.

While duration is the primary acoustic correlate of vowel length, other phonetic factors may also play a role in the perception of long vowels as opposed to short ones. Pitch is one such factor (Kinoshita, Behne, and Arai 2002).[2] Since pitch changes may occur in long vowels but not in short ones in Tokyo Japanese, the presence of a pitch fall within a vowel signals that the vowel is phonologically long. This can be illustrated by the minimal pair of words in (8) which are accented on their initial syllables: Perceptually, pitch falls within the syllable /bii/ in (8a), whereas it falls between the two syllables, /bi/ and /ru/, in (8b).

(8) a. bii.ru 'beer'
b. bi.ru 'building'

Finally, modern Japanese has some diphthongs in addition to short and long vowels. There is some dispute in the literature as to which vowel sequence constitutes a diphthong as opposed to hiatus, or vowel sequences across a syllable

1 It also adopts word-final schwas with a r-coloring (spelt as *–er, -ar, -or, -ur, -ir*) as long vowels, which accounts for the high frequency of word-final /aa/ in loanwords.

2 Hirata and Tsukada (2009) also show that formants are more widely dispersed in long vowels.

boundary. Phonological considerations suggest that only three vowel sequences function as diphthongs in the language, i.e. /ai/, /oi/ and /ui/ (Kubozono 2004, 2005, 2008c). The fact that all these diphthongs end in /i/ is not a coincidence since vowel sequences ending in /u/ such as /au/, /ou/, /eu/ and /iu/ underwent a historical change in native and SJ words whereby they turned into long vowels (see Kubozono Ch. 5, this volume, for a detailed analysis of diphthongs and vowel coalescence). Some examples are given below.

(9) /kjau/ → /kjoo/ ‘capital’ (京)
/koukou/ → /koo.koo/ ‘high school’ (高校)
/teuteu/ → /tjoo.tjo(o)/ ‘butterfly’ (蝶々)
/riu/ → /rjuu/ ‘dragon’ (竜)
/iu/ → /juu/ ‘to say’ (言う)

3 Consonant inventory

3.1 Phonemes and allophones

The inventory of consonantal phonemes in modern Japanese is given below (adapted from Shibatani 1990: 159). See section 8 below for the phonetic symbols adopted in this volume.

Table 1: Consonant system of modern Japanese

	labial	dental-alveolar	palatal	velar	glottal
plosive	p	t		k	
	b	d		g	
fricative		s			h
		z			
nasal	m	n			
liquid			r		
glide	w		j		

This system might look simpler than the consonant systems of other languages. This is not quite true, however, since Japanese has many more consonants if allophones are also considered. For example, /s/ has two allophones that are distributed in a complementary fashion in native words: [ʃ][3] appears before /i/ and [s] before other vowels. Likewise, /t/ is realized in three forms: the affricate [tʃ] before /i/, the

3 While the symbol [ɕ] is often used for this consonant in the literature, [ʃ] is used in this volume. See section 8 below.

affricate [ts] before /u/, and the dental stop [t] anywhere else. Similarly, /h/ has three allophonic forms: the palatal fricative [ç] before /i/, bilabial fricative [ɸ] before /u/, and glottal fricative [h] before any other vowel.

While these 'allophones' are thus in complementary distribution in the native vocabulary, many of them are not complementary in SJ words and loanwords. In SJ words, for example, [ʃ] and [s] appear in the same context: e.g. [ʃakɯ] 'to serve *sake*' vs. [sakɯ] 'a fence'. [tʃ] and [t] are not complementary, either: e.g. [tʃa] 'tea' vs. [ta] 'others'. [ç] and [h] also appear before the same vowel: [çakɯ] 'hundred' vs. [hakɯ] 'beat'. Each of these consonants can be said to have established itself as an independent phoneme in the consonant system of modern Japanese.

This tendency has been accelerated recently as loanwords from English and other languages introduced new sequences of sounds into Japanese (see the discussion in Pintér, this volume, for details). For example, [t] now appears not only before [a] but before [i] and [ɯ], too: [ti:] 'tea' [ti:ʃatsɯ] 'T shirt'; [tɯ:] 'two', [tɯ:rɯ:zɯ] 'Toulouse'. Similarly, [ts] appears before [a] as well as before [ɯ]: e.g. [mo:tsarɯto] 'Mozart', [tsɯa:] 'tour'. Moreover, [ɸ] may be combined with vowels other than [ɯ], creating CV sequences contrasting with [ɸɯ]. This is shown in (10).

(10) a. [ɸan] 'fan', [ɸaito] 'fight'
b. [ɸitto] 'fit', [ɸi:ringɯ] 'feeling'
c. [ɸensɯ] 'fence', [ɸe:sɯ] 'face'
d. [ɸo:rɯ] 'fall', [ɸonto] 'font'

Recent loanwords have also 'filled' the native phonotactic gaps in the distribution of some consonant phonemes. For example, /w/ has been allowed to combine only with /a/ in the phonology of native and SJ words, as evidenced by the alternations such as /suwaru/ 'to sit' and /sueru/ 'to set', the latter deriving from /suweru/ historically via /w/ deletion. This situation is being changed by the introduction of some recent loanwords such as [wetto] 'wet', [wo:ta:] 'water', [wisuki:] 'whisky', [witto] 'wit', and [winta:] 'winter'.

A similar situation can be found with [j], which combined only with back vowels (/a/, /o/ and /u/) in the traditional phonological system. This consonant can now be combined with /e/ in some loanwords, expanding the phonological context where it can appear: e.g. [jeritsin] 'Yeltsin (former Russian president)' (Ito and Mester 1995).

3.2 Voice in consonants

As in many other languages, consonants in modern Japanese exhibit a contrast in voice, but not all consonants participate in this contrast. Voiceless consonants have their voiced counterparts in the system, but not vice versa. In order to understand this point, we should first note that voiced consonants fall into two groups, voiced

obstruents and non-obstruents (or sonorants), of which only the first group involves a phonological contrast with voiceless consonants in modern Japanese. Thus, voiced stops ([b], [d], [g]) and fricative ([z]) contrast with voiceless ones ([p], [t], [k] and [s]). On the other hand, voiced non-obstruents ([m], [n], [r], [j] and [w]) do not have voiceless counterparts at least at the phonemic level.

In natural languages, voiceless obstruents are supposed to be unmarked as against voiced obstruents. For one thing, every language that has voiced obstruents is likely to have voiceless ones, although not vice versa. Maddieson (1984: 27) reports that "a language with only one stop series almost invariably has plain voiceless plosives." Similarly, Yavaş (1998: 173) mentions that "the existence of the voiced obstruents implies the existence of its voiceless counterparts." In phonological development, too, children acquire voiced obstruents only after they have acquired voiceless ones: "Children will ... use voiceless unaspirated stops before acquiring the pattern of voicing types that is contrastive in their language" (Macken 1980: 163).

On the other hand, non-obstruents such as nasals, liquids and semivowels are typically voiced. For example, voiced nasals ([m], [n]) are unmarked as against voiceless nasals ([m̥], [n̥]) in natural languages, as can be seen from the fact that a diacritic symbol is added in the phonetic descriptions of the latter. Seen in this light, the distribution of consonants in the Japanese consonant system looks quite natural: It contains unmarked sounds (voiceless obstruents and voiced non-obstruents) in its core part, plus some marked sounds (voiced obstruents). In the traditional Japanese linguistics, the former group is called "seion" (清音), or pure sounds, whereas the latter group is called "dakuon" (濁音), or impure sounds. This grouping is summarized in Table 2, where the shaded part denotes "seion", and the others are "dakuon".[4]

Table 2: Two-way classification of consonantal phonemes in modern Japanese

	voiceless	voiced
obstruents	p, t, k, s, h	b, d, g, z
non-obstruents	–	m, n, r, w, j

Voiced obstruents did not exist in the consonant system of old Japanese, at least in word-initial position. They developed in the language in several ways (see Vance, this volume, and Takayama, this volume, for the historical development of voiced obstruents in Japanese). One major source is SJ words, or old loanwords from Chinese, which had many voiced obstruents in morpheme-initial positions. This can be understood by the comparison between the "on-yomi" (SJ pronunciation) and "kun-yomi" (native pronunciation) of Chinese characters.

4 /p/ is not a *seion* in the strict sense of the term, but is often called "han-dakuon" (half *dakuon*). It is a consonant introduced primarily by loanwords.

Table 3: Comparison of *on-yomi* and *kun-yomi*

Chinese character	*on-yomi*	*kun-yomi*	Gloss
外	gai	so.to	outside
人	zin	hi.to	man
男	dan	o.to.ko	male
女	zyo	on.na	female

Another major source of voiced obstruents is the sound change known as *rendaku*, or sequential voicing, in native words. This voicing process turned voiceless obstruents into their voiced counterparts in the initial syllable of non-initial members of compounds. Some examples are given below (see Ito and Mester 2003 and Vance, this volume, for a full discussion of this process).

(11) a. osiroi 'powder' + hana 'flower' → osiroi-bana
'a tropical American flower, also known as a four-o'clock'
b. to 'door' + tana 'shelf' → to-dana 'closet, cupboard'
c. hira 'flat' + kana 'kana syllabary' → hira-gana 'hiragana syllabary'
d. nobori 'upward' + saka 'slope' → nobori-zaka 'uphill slope'

What is interesting about *rendaku* is that /b/ alternates with /h/ rather than /p/. This can be attributed to a change by which /p/ turned into /h/ (phonetically, [ç], [ɸ], [h]) in word-initial position in the course of the history. This historical change shows its trace in the alternation between /h/ and /p/ in the morphophonology of modern Japanese. For example, the word /hi.jo.ko/ 'a baby bird' can be related with the mimetic expression /pi.jo.pi.jo/ which denotes the sound produced by baby birds. Likewise, the noun /hi.ka.ri/ 'light' is historically related with the mimetic expression /pikari/ which describes lightening. In these words, /p/ turned into /h/ in ordinary nouns, but remained unchanged in onomatopoeic expressions.

3.3 Moraic obstruents and nasals

Among consonants in Japanese, so-called moraic obstruents and nasals are different from other consonants in many ways. These are consonantal elements that are called "sokuon" (促音) and "hatsuon" (撥音), and in the traditional literature are often represented as /Q/ and /N/, respectively. These two types of consonants are similar to each other in the following respects. First, they only occur in the coda position of the syllable. Second, as moraic consonants, they contribute to the weight of the syllable in Tokyo Japanese so that syllables containing these consonants in the coda position are counted as bimoraic as opposed to monomoraic. Third, both

of them are phonetically homorganic with the following consonant: they share the place of articulation with the onset consonant of the following syllable. This is illustrated in (12) and (13), respectively, where the traditional archiphonemic symbols /Q/ and /N/ are used for the sake of description.

(12) a. /iQ.pai/ [ippai] 'one cup' (一杯), /raQ.pa/ [rappa] 'trumpet', /taQ.paa/ [tappa:] 'tupper'
b. /iQ.tai/ [ittai] 'one body' (一体), /baQ.ta.ri/ [battari] 'with a thud, suddenly', /baQ.taa/ [batta:] 'batter'
c. /iQ.kai/ [ikkai] 'one time' (一回), /gaQ.ka.ri/ [gakkari] 'with a disappointment', /saQ.kaa/ [sakka:] 'soccer'

(13) a. /aN.ma/ [amma] 'massage', /aN.ba/ [amba] 'pommel horse', /raN.pu/ [rampɯ] 'lamp'
b. /aN.na/ [anna] 'Anna (girl's name), /aN.da/ [anda] 'a hit in baseball', /saN.ta/ [santa] 'Santa Clause'
c. /maN.ga/ [maŋga] 'cartoon, anime', /taN.ka/ [taŋka] '*tanka* poem', /taN.ku/ [taŋkɯ] 'tank'

In semantic terms, moraic obstruents and nasals are similar to each other in that they are often used as a kind of infix for emphasis. When used for emphasis, they are complementary with each other: moraic obstruents appear before voiceless consonants, whereas moraic nasals appear elsewhere (Kuroda 1965). This is shown in (14), where /ma/ is an emphatic prefix often combined with the moraic consonants.

(14) a. maQ.pu.ta.tu 'in half' (< hu.ta.tu 'two')
maQ.pi.ru.ma 'mid-day' (< hi.ru.ma 'daytime')
maQ.ku.ro 'deep black' (< ku.ro 'black')
maQ.si.ro 'pure white' (< si.ro 'white')
b. maN.ma.ru 'perfect circle' (< ma.ru 'circle')
maN.na.ka 'the very center' (< na.ka 'middle, center')
suN.goi 'fantastic!' (< su.goi 'dreadful, wonderful')
aN.ma.ri '(not) too much' (< a.ma.ri '(not) very')

While moraic obstruents and nasals are similar to each other in these respects, they differ in other ways. For one thing, moraic obstruents cannot appear in word-final position, except in some interjectives such as /aQ/ [aʔ]~[atʔ] 'Oh, Oh dear' and /eQ/ [eʔ]~[etʔ] 'really?'. In fact, they form a geminate obstruent together with the onset consonant of the following syllable: e.g. [makkɯro] 'deep black'. In contrast, moraic nasals can appear freely in word-final position: e.g. /en/ 'yen, the Japanese currency', /kan/ 'tube', /sen/ 'line'. Phonetically, these word-final nasals are often

described as nasal(ized) vowels or the like (Kuroda 1965; Kawakami 1977). Kawahara (Ch. 1, this volume) and Kawagoe (this volume) provide a more detailed description of moraic obstruents.

4 Mora and syllable

4.1 Mora versus syllable

The distinction between the mora and the syllable is crucial in Japanese phonetics and phonology. Since these two units are used in most chapters of this volume, we would like to discuss them in some detail here. In native words, they overlap in most cases. For example, /na.go.ja/ 'Nagoya' and /to.jo.ta/ 'Toyota' both consist of three syllables and three moras: each mora corresponds to a syllable. However, these two phonological units often fail to overlap in many SJ words and loanwords. The words /nip.pon/ 'Japan' and /zja.pan/ 'Japan' are both made up of two syllables, but they contain four and three moras, respectively.

The discrepancy between the mora and the syllable arises because some moras cannot constitute a syllable on their own. In other words, moras fall into two types, those that can constitute a syllable on their own and those that are always attached to another mora to form a syllable. The former is labeled as "jiritsu haku" (自立拍, independent mora), and the latter as "tokushu haku" (特殊拍, meaning 'special mora'), "huzoku haku" (付属拍, or dependent mora) or "moora onso" (moraic phonemes) in the traditional literature (Kawakami 1977). In the more recent literature, they are also called "syllabic mora" vs. "non-syllabic mora" (Kubozono 1989, 1999a), or "head mora" vs. "non-head mora" (Ito and Mester 1993; Kubozono 2012a,b).

Those moras that cannot constitute a syllable on their own fall into four types in Tokyo Japanese: (a) the second half of long vowels, (b) the second half of diphthongs (ai, oi, ui), (c) moraic nasals, or the coda nasals, and (d) moraic obstruents, or the first half of geminate consonants. These are often represented as /R/, /J/, /N/ and /Q/ in traditional descriptions. These moras form a syllable together with their preceding moras, as shown below.

(15) a. too 'ten, tower', koo.to 'coat, court', nii.san 'elder brother', o.too.to 'younger brother', kaa.ten 'curtain'
b. gai.ko.ku 'foreign country', kan.sai 'Kansai area', sai.daa 'cider (lemonade)'
c. son 'loss', san 'three', ton.da 'jump (past tense)', ron.don 'London'
d. nip.pon 'Japan', hat.ten 'development', ro.ket.to 'rocket', gak.koo 'school'

The following table shows the discrepancy between mora and syllable counts in some proper names. Dots and hyphens indicate syllable and mora boundaries, respectively.

Table 4: Syllable count versus mora count

Word	Gloss	Syllable count	Mora count
toyota	Toyota	3 (to.yo.ta)	3 (to-yo-ta)
nissan	Nissan	2 (nis.san)	4 (ni-s-sa-n)
honda	Honda	2 (hon.da)	3 (ho-n-da)
kurinton	Clinton	3 (ku.rin.ton)	5 (ku-ri-n-to-n)
bussyu	Bush	2 (bus.syu)	3 (bu-s-syu)
obama	Obama	3 (o.ba.ma)	3 (o-ba-ma)

Two points must be noted about the four types of 'special moras' in (15). First, not all languages count these elements as 'moraic', or as elements contributing to syllable weight. Some languages allow only long vowels and diphthongs in (15a,b) to count as two moras, some give a moraic status to coda nasals in (15c) as well, and others count all elements in (15) including coda obstruents in (15d) as moraic (Zec 1995). Japanese belongs to the last type of language in this classification.

What is relevant here is the existence of a hierarchy or implicational law as shown in (16), which indicates that coda obstruents can have a moraic status only in a system in which all other elements have a moraic status. It also indicates that coda nasals can count as one mora only when the second half of diphthongs and long vowels is counted as one mora (Zec 1995). Seen conversely, this hierarchy precludes the possibility of a system in which coda obstruents but not coda nasals count as one mora as well as a system where coda nasals and obstruents are moraic but the second half of diphthongs and long vowels are not.

(16) a. System A: second half of diphthongs and long vowels
b. System B: second half of diphthongs and long vowels__coda nasals
c. System C: second half of diphthongs and long vowels__
coda nasals__coda obstruents

Another point to note about (15) is that the four types of special moras do not appear equally frequently in Japanese. For example, moraic nasals occur twice as frequently as moraic obstruents if counted by type frequency (Hayashi 1982: 319). There seems to be no statistical study that compared their relative frequency with that of long vowels and diphthongs or the frequency of long vowels vs. diphthongs.

4.2 Syllable weight

As is clear from the foregoing discussion, there are two types of syllables: (i) those that are made up of one mora, i.e. a syllabic (independent) mora, and (ii) those that consist of two moras, a syllabic mora plus a non-syllabic (special) mora. (The possibility of trimoraic syllables will be discussed shortly). These two types of syllables are called ‘light’ (or ‘short’) syllables and ‘heavy’ (or ‘long’) syllables. By definition, light syllables are monomoraic and heavy syllables are bimoraic.

It is generally assumed that there was no contrast in syllable weight in Old Japanese. As long vowels, diphthongs and coda consonants developed and acquired a moraic status in the history of the language, it has become necessary to distinguish between independent and special moras, or equivalently, between light and heavy syllables. Not surprisingly, light syllables are much more common than heavy ones in modern Japanese. According to Kubozono (1985), the ratio between light and heavy syllables is about two to one (2:1).

While both monomoraic and bimoraic syllables are popular, trimoraic syllables are not. These syllables, often called ‘superheavy’ or ‘overlong’ syllables, may occur only in loanwords, while they do not generally occur in native and SJ words due to various phonotactic and morphophonological constraints to be discussed in section 6 below. Superheavy syllables will typically consist of a long vowel or diphthong followed by a coda consonant, as exemplified below.

(17) wain ‘wine’, rain ‘line, Rhine’, taun ‘town’,
su.pein ‘Spain’, rin.kaan ‘Lincoln’, gu.riin ‘green’

A careful phonological analysis suggests, however, that these seemingly trimoraic syllables actually consist of two syllables rather than one. One piece of evidence to support this view comes from the accent analysis of compound nouns of which the words in (17) form the first member. The basis of the compound noun accent rule in Tokyo Japanese is to place a compound accent on the final syllable of the first member if the second member is one or two moras long. If applied to the compound nouns just mentioned, this accent rule generally places an accent on the second mora of the quasi-trimoraic syllables, suggesting that there is a syllable boundary between this mora and the preceding mora (Kubozono 1999a).[5] In this and subsequent chapters in this volume, word accent is denoted by an apostrophe (’), which indicates the position of an abrupt pitch fall in Tokyo Japanese.[6]

5 A similar observation can be made of non-compound loanwords like /a.na.u’n.su/ ‘announce’ and /a.na.u’n.saa/ ‘announcer, or news reader’. The position of the accent in these words suggests that /naun/ consists of two syllables, /na/ and /un/.

6 In some Japanese dialects, pitch rise rather than pitch fall is distinctive. See Uwano (2012) for details.

(18) rain + ka.wa → ra.i'n.ga.wa, *ra'in.ga.wa 'River Rhine'
taun + si → ta.u'n.si, *ta'un.si 'town magazine'
su.pein + ka.ze → su.pe.i'n.ka.ze, *su.pe'in.ka.ze 'Spanish influenza'
rin.kaan + hai → rin.ka.a'n.hai, *rin.ka'an.hai 'Lincoln Cup'

That the quasi-trimoraic strings actually consist of two syllables can be shown more clearly if the data of Kagoshima Japanese are considered. Unlike Tokyo Japanese, this dialect is sensitive to syllable boundaries and its accent rule places a high pitch on the penultimate syllable in most loanwords as well as compound nouns whose first member is a loanword. An accent test of this dialect shows that there is a syllable boundary between the first and second moras of the three-mora strings in question (Kubozono 2004).

(19) ta.UN.si 'town magaine'
su.PE.in 'Spain'
su.pe.IN.zin 'Spanish people'
rin.ka.AN.hai 'Lincoln Cup'

The idea that superheavy syllables are disfavored and tend to be avoided in modern Japanese can further be reinforced by several other independent pieces of evidence. First, so-called pre-nasal vowel shortening (Lovins 1975) produces a bimoraic syllable out of a sequence of segments that would otherwise be borrowed as a trimoraic syllable in Japanese (Kubozono 1995a, 1999a). This is exemplified in (20), where English forms are given in the input[7] and the loanword forms in Japanese in the output. The relevant portion is underlined.

(20) a. [faun.dei.ʃən] → [ɸan.deː.ʃon] 'foundation'
b. [keim.bridʒ] → [ken.bu.rid.dʒi] 'Cambridge'
c. [steind glæs] → [su.ten.do.gu.ra.sɯ] 'stained glass'
d. [griːn.piːz] → [gu.rin.piː.sɯ] 'green peas'
e. [kɔːnd biːf] → [kon.bii.ɸɯ] 'corned beef'

Generally speaking, [n] in the coda in English words is borrowed as a moraic nasal. Moreover, tense vowels and diphthongs in English are borrowed as long vowels and diphthongs in Japanese. One notable exception to the latter rule is the tense vowels before a nasal consonant. As the examples in (20) show, English tense vowels in this context tend to be borrowed as short vowels. Thus, in (20d), the same tense vowel /iː/ [iː] in English turned into a short vowel in 'green' but into a long

7 To show the moraic structure of English words explicitly, tense vowels are represented with a length marker [ː] here.

vowel in ‘peas’ in the word ‘green peas’. The rule of pre-nasal vowel shortening permits a certain number of exceptions as shown in (17), but it is clear that this rule has a function of turning potential trimoraic syllables into bimoraic ones via vowel shortening.

Another process that has prevented Japanese from creating superheavy syllables is antigemination, whereby consonant gemination is blocked under certain environments. Consonant gemination is a popular process in the loanword phonology of Japanese that geminates voiceless obstruents in the coda position of English words as exemplified in (21) (Kawagoe and Arai 2002; Kawagoe and Takemura 2013; Kubozono, Takeyasu, and Giriko 2013; see also Kawagoe, this volume). Seen from the viewpoint of syllable weight, this process creates a sequence of heavy and light syllables in the output.

(21) pak.ku ‘pack’
kap.pu ‘cup’
kat.to ‘cut’
bat.to ‘bat’

This process is subject to various restrictions as discussed in detail by Kawagoe (this volume). One such restriction concerns the length of the preceding vowel. Namely, the process is blocked if the vowel before the voiceless obstruent is a tense one or a diphthong in the source language. These tense vowels and diphthongs are borrowed as long vowels in Japanese and block the gemination process. Some examples are given in (22).

(22) paa.ku, *paak.ku ‘park’
kaa.pu, *kaap.pu ‘carp’
kaa.to, *kaat.to ‘cart’
kai.to, *kait.to ‘kite’

Pre-nasal vowel shortening in (20) and antigemination in (22) may be described by phonotactic constraints that prohibit long vowels and diphthongs from occurring with the following moraic nasal or obstruent. These phonotactic constraints do not explain the general nature of the phenomena, however. What is truly crucial here is that both processes contribute to avoiding the creation of trimoraic syllables in the output, a prosodic structure that is disfavored across languages (Árnason 1980; Sherer 1994; Zec 1995). Seen in this light, it is clear that they have much in common with the syllabification fact revealed by accent phenomena in (18) and (19). All these phenomena conspire to avoid creating superheavy syllables in Japanese (see Kubozono Ch. 8, this volume, for some potential exceptions such as the native word /to’otta/ ‘passed (past tense of *pass*)’).

In sum, the syllable structure of the language can be described as (Cj)V(VC), where materials in the parentheses are optional.[8] /V/ and /C/ in the second parentheses cannot generally co-occur with each other due to the strict constraint banning superheavy syllables.

4.3 Mora's roles

The roles of moras in Japanese phonetics and phonology have been widely discussed in the literature: see the series of work by Kubozono (Kubozono 1985, 1989, 1999a) and Otake (Otake et al. 1993), for example (see also Otake, this volume, for a full summary). In order to argue for the relevance of this prosodic unit, it is necessary to demonstrate that one and the same phenomenon can be better generalized by the mora than by the syllable. Specifically, it must be shown that bimoraic syllables, i.e. heavy syllables, behave similarly to a sequence of two monomoraic syllables, i.e. (light + light) sequence, and differently from monomoraic monosyllables.

The most famous role of the mora is probably its role as a counting unit in the meter of traditional poems such as *tanka, haiku* and *senryū* (Bekku 1977; Homma 1984). *Tanka* poems, literally meaning 'short poem', for example, consist of five lines of which the first and three lines are made up of five moras and the remaining lines, seven moras (5-7-5-7-7). *Haiku* and *senryū* are shorter versions consisting of three lines of which the first and third lines are made up of five moras (5-7-5). To take one example from Kobayashi Issa's famous poetry, the second line of the *haiku* below has seven units by mora count, but six units by syllable count: the word /is.sa/ 'Issa, the composer's name' is disyllabic and trimoraic (hyphens indicate mora boundaries).

(23)	ya-se-ga-e-ru	'a skinny frog!'
	ma-ke-ru-na i-s-sa	'don't give up. Issa'
	ko-re-ni-a-ri	'is here'

Likewise, the following *senryū* composed by an elementary school child consists of five, seven and five units by mora count, but three, five and four units by syllable count. This example as well as Issa's *haiku* above show that it is the mora and not the syllable that is used to define the meter of Japanese poetry.

(24)	ni-ho-n-zyu-u	'all over Japan'
	a-t-ti ko-t-ti-de	'here and there'
	ta-ma-go-t-ti	'(everyone has) a tamagotchi toy'

8 /j/ is a glide that can appear between a consonant and a back vowel. /CjV/ is a structure popular in SJ words and loanwords: /tjuu/ [ʧɯː] 'middle, average', /mjoo/ [mjoː] 'strange', /sjatu/ [ʃatsɯ] 'shirt'. See Pintér (this volume) for more data and analysis.

The mora's role has been discussed in the description of speech rhythm, too. Thus, the concept of "mora timing" comes from the idea that the mora is a timing unit that repeats itself isochronously in natural speech. This idea has been discussed in much detail in the literature, sometimes supported and sometimes refuted (or doubted) by experimental evidence (see Warner and Arai 2001 for a summary). It is certainly true that moras in Japanese do not have equal phonetic durations since they vary durations considerably due to segmental, prosodic and other factors. However, it is also true that word durations can be more or less predicted in the language by the number of moras contained in the word (Port, Dalby, and O'Dell 1987).

The mora has been used in the description of phonological phenomena, too, where the concept of "mora counting" is widely known, particularly in the analyses of word accent (McCawley 1968). This concept is based upon the idea that the position of word accent is determined by counting the number of moras in the word; in other words, the mora is used as a unit to measure phonological distances in the computation of word accent. For example, loanwords are supposed to be subject to the antepenultimate mora rule that places the accent on the third mora from the end (McCawley 1968): e.g. /su.to'.re.su/ 'stress', /su.to'.roo/ 'straw', /ku.ri.su'.ma.su/ 'Christmas'. The mora-counting priniciple applies to many other accent rules of Tokyo Japanese, which prompted McCawley (1978) to define the dialect as a 'mora-counting' language.

The mora plays a crucial role as a counting unit in other phonological processes, too. For example, it is known that *rendaku* is often constrained by the phonological length of the first member of compound nouns.[9] Thus, the noun /hon/ 'book' undergoes this voicing process if it is attached to a three-mora or longer noun, as in (25a), but generally not if the preceding element is either monomoraic or bimoraic, as in (25b) (Ohno 2000).

(25) a. man.ga + hon → man.ga-bon 'a comic book'
mi.do.ri + hon → mi.do.ri-bon 'green book'
et.ti + hon → et.ti-bon 'an erotic book'

b. a.ka + hon → a.ka-hon 'red book, a textbook for entrance exams'
e.ro + hon → ero-hon 'an erotic book'
e + hon → e-hon 'a picture book'

Morphological processes are also sensitive to the mora. For example, compound truncation generally combines the initial two moras of the two elements of the compound noun (Ito and Mester 2003; Kubozono 1999a,b).

9 According to Rosen (2003), *rendaku* is generally sensitive both to the number of moras in the first element and to the number of moras in the second element. See Vance (this volume) for details.

(26) a. ki.mu.ra ta.ku.ya → ki.mu-ta.ku ‘Kimura Takuya, a famous singer/actor’
b. po.ket.to mon.su.taa → po.ke-mon ‘pocket monster, or Pokémon’
c. han.gaa su.to.rai.ki → han-su.to ‘hunger strike’
d. han.bun don.ta.ku → han.don ‘half, Zontag; a half day off’

We have seen that the mora is an indispensable unit in a wide range of phenomena of Japanese, phonetic, phonological and morphological. What is important in considering the roles of the mora in these processes is that the same generalizations cannot be made on the basis of the syllable. More specifically, one must demonstrate that heavy syllables behave differently from light syllables yet similarly to a sequence of two light syllables. To take one example from (25), a crucial case is the contrastive *rendaku* behaviors of /et.ti/ and /e.ro/, which share morphological, semantic and syllable structures but differ in the number of moras involved: /et.ti/ has three moras and patterns with other three-mora nouns like /mi.do.ri/ ‘green’, whereas /e.ro/ consists of two moras and does not trigger *rendaku* similar to other bimoraic nouns.

Similarly, the outputs of the truncation process described in (26) are all made up of four moras which vary in the composition of syllables: four light syllables in (26a), two light syllables followed by a heavy syllable in (26b), a heavy syllable followed by two light syllables in (26c), and two heavy syllables in (26d). The uniformity of the output forms cannot be captured if their phonological length is measured by the syllable.

The roles of the mora in (26) and other phonological processes have led to the notion of ‘bimoraic foot’, which assumes that a sequence of two moras forms a prosodic unit above the mora in the prosodic hierarchy. This unit was first posited for the description of traditional poems (Bekku 1977), while it has been shown to be indispensable for the description of many linguistic phenomena, too. The pioneering work in the latter field is Poser (1990), whose analysis has been extended to various linguistic phenomena: loanword accent (Katayama 1998; Shinohara 2000), compound accent (Kubozono 1995b, 1997; Kubozono, Ito, and Mester 1997), compound truncation (Ito and Mester 2003) and the formation of *zuuzya-go (zuuja-go)*, a jazz musicians’ secret language (Tateishi 1989; Ito, Kitagawa, and Mester 1996). It is also instrumental in explaining why monomoraic outputs are disfavored and tend to be lengthened in phrases. Vowel lengthening in monomoraic nouns is illustrated below (Ito 1990; Mori 2002).

(27) a. /go/ ‘five’ → [go:] as in /goo.i.ti.goo zi.ken/ ‘5.15 incident’
b. /ni/ ‘two’ → [ni:] as in /nii.nii.ro.ku zi.ken/ ‘2.26 incident’
c. /ne/ → [ne:] ‘the year of the Rat (in the Chinese zodiac)’

4.4 Syllable's roles

While various arguments can be found in the literature in favor of the mora, relatively little has been discussed about the relevance of the syllable in Japanese. This can be attributed in large part to the traditional typology by Trubetzkoy (1958) and Kindaichi (1967), who claimed that the syllable and the mora cannot coexist in a single prosodic unit. They classified languages into two types: "mora" (or mora-based) languages like Tokyo Japanese and "syllable" (syllable-based) languages like English. A similar idea can be found in Sibata's (1962) classification of Japanese dialects into "mora dialects" and "syllabeme dialects".

This rigid dichotomy has been challenged in the past few decades. One the one hand, it has been demonstrated that the distinction between light and heavy syllables plays a pivotal role in the phonology of English and many other languages for, for example, computing stress patterns and explaining compensatory lengthening (see Hayes 1989, 1995). Since the idea of syllable weight hinges upon the notion of mora, these analyses have integrated the mora as a basic unit in the description of what was formerly described as syllable(-based) languages.

On the other hand, the syllable has been recognized as a relevant phonological unit in the description of mora languages. In Tokyo Japanese, the syllable plays somewhat subsidiary roles, but there are many phenomena in which the bimoraic monosyllables pattern with monomoraic monosyllables and not with bimoraic disyllabic words. In loanword accentuation, for example, the accent falls on the fourth mora from the end of the word if the antepenultimate mora is a special mora, or the second mora of a heavy syllable.

(28) ro'n.don, *ron'.don 'London'
wa.si'n.ton, *wa.sin'.ton 'Washington'
sa'i.daa, *sai'.daa 'cider, or lemonade'

What is crucial here is the fact the accent apparently 'shifts' one mora to the left, and not one mora to the right. This indicates that the accent is placed on the first mora of the heavy syllable if its second mora is the designated as an accent position by accent rules. This led McCawley (1968) to propose the following accent rule for loanwords, which actually accounts for the accentuation of *accented* native and SJ words, too (Kubozono 2006, 2008a,b, 2011, 2013).[10] Since the syllable is the unit bearing the accent in Tokyo Japanese, it is labeled as a 'mora-counting, syllable language' (McCawley 1978).

(29) Accent falls on the syllable containing the antepenultimate mora.

---

10 The antepenultimate effect can be attributed to the interaction of several general constraints, particularly nonfinality and edgemostness (Katayama 1998; Kubozono 2008a,b, 2011). Under this analysis, the antepenultimate effect is an epiphenomenon that does not exist as an independent principle or rule.

Some accent rules are sensitive to both the mora and the syllable in measuring phonological length. A good example is the rule responsible for the accentuation of personal names whose second member is /ta.roo/ ‘Taro’ (Kubozono 1999a,b). These names exhibit three accent patterns depending on the phonological length of the first member: (i) the unaccented pattern if the first member is monosyllabic, (ii) a pattern with an accent on the last syllable of the first member if the first member is disyllabic and bimoraic, and (iii) a pattern with an accent on the initial syllable of /ta.roo/ if the first member is longer than two moras. These three patterns are illustrated below. Of these, the third pattern represents the general accent pattern of compound nouns with a three-mora second member, e.g. /ti.ka.ra-si’.go.to/ ‘power, work; heavy labor’, /bii.ti-ba’.ree/ ‘beach, volleyball; beach volleyball’.

(30) a. ko-ta.roo ‘Ko-taro’
kin-ta.roo ‘Kin-taro’
koo-ta.roo ‘Koo-taro’
b. mo.mo’-ta.roo ‘Momo-taro’
a.ma’-ta.roo ‘Ama-taro’
i.ti’-ta.roo
c. ti.ka.ra-ta’.roo ‘Chikara-taro’
ka.ree-ta’.roo ‘Curry-taro’
u.ru.to.ra.man-ta’.roo ‘Ultra-man Taro’

The boundary between (30a) and (30b) is defined by the number of syllables in the first member, i.e. monosyllabic vs. disyllabic, whereas the boundary between (30b) and (30c) is determined by the number of moras, i.e. bimoraic vs. trimoraic. It is interesting that one and the same rule employs the two prosodic units to produce different accent patterns.

The syllable plays an indispensable role in other accent rules, too (Kubozono 1999a,b). Four-mora loanwords, for example, tend to be unaccented, i.e. lack an abrupt pitch fall in the phonetic output, if they end in a sequence of two light syllables: e.g. /kon.so.me/ ‘consommé’, /mo.na.ri.za/ ‘Mona Lisa’, /ai.o.wa/ ‘Iowa’, /a.ri.zo.na/ ‘Arizona’, /mo.su.ku.wa/ ‘Moscow’ (Kubozono 1996). On the other hand, four-mora loanwords of other syllable structures disfavor the unaccented pattern: e.g. /pa.re’e.do/ ‘parade’, /o.re’n.zi/ ‘orange’, /o.ha’i.o/ ‘Ohio’, /ro’n.don/ ‘London’, /su.to’.roo/ ‘straw’. Table 5 shows the extent to which the accentuation of four-mora loanwords depends on their syllable composition (Kubozono 1996, 1999a,b, 2006): ‘L’ and ‘H’ stand for light and heavy syllables, respectively. That accent patterns vary greatly within four-mora words can be accounted for if and only if they are analyzed in terms of the syllable.

Table 5: The ratios of unaccented words in four-mora loanwords as a function of syllable structure

Syllable composition	LLLL	HLL	LHL	LLH	HH
Ratio of unaccentedness	54%	45%	24%	19%	7%
Example (unaccented)	a.ri.zo.na	mai.na.su	be.ran.da	pe.ri.kan	ai.ron
	'Arizona'	'minus'	'veranda'	'pelican'	'iron'
Example (accented)	to.ra'.bu.ru	ma'i.ru.do	bu.ra'n.ko	su.to'.roo	na'i.ron
	'trouble'	'mild'	'swing'	'straw'	'nylon'

Looking beyond word accent, the syllable has been shown to be indispensable for the description of many morphological processes, too. In the truncation of loanwords, for example, not only monomoraic forms but also bimoraic forms consisting of one syllable are banned (Ito 1990).[11] In other words, bimoraic forms are acceptable if they consist of two syllables, e.g. /su.to/ 'strike', but not if they consist of one syllable, as the following examples show. The fact that bimoraic words display contrastive behaviors depending on their syllabic composition indicates that the syllable plays a critical role.

(31) roo.tee.syon → roo.te, *roo 'rotation'
paa.ma.nen.to (wee.bu) → paa.ma, *paa 'permanent wave'
pan.fu.ret.to → pan.fu, *pan 'pamphlet'
don.ki.hoo.te → don.ki, *don 'Don Quijote Group, a Japanese chain store'

Moreover, three-mora outputs consisting of a light syllable followed by a heavy syllable are also illicit. This contrasts with other types of trimoraic outputs that are well-formed. (32) compares these ill-formed and well-formed outputs.

(32) a. ro.kee.syon → ro.ke, *ro.kee 'location'
de.mon.su.to.ree.syon → de.mo, *de.mon 'demonstration'
b. te.re.bi.zyon → te.re.bi 'TV'
pan.fu.ret.to → pan.fu 'pamphlet'
paa.ma.nen.to (wee.bu) → paa.ma 'permanent wave'

The tendency to avoid creating light-heavy disyllables is observed in a wide range of phenomena in Japanese including baby talk words, *zuuzya-go* formation and historical changes (Kubozono 2000, 2003). Overall, the syllable as well as the

11 The same constraint is at play in the truncation of native words, too. Truncation of adjectives, for example, combines the initial two moras of the stem with the adjectival marker /i/: e.g. /mu.zu.ka.si-i/ → /mu.zu-i/ 'difficult', /ki.sjo.ku.wa.ru-i/ → /ki.sjo-i/ 'disgusting'. However, the initial three moras must be taken from the stem if the initial two moras form a heavy syllable: /ut.to.o.si-i/ → /ut.to-i/, */ut-i/ 'gloomy'.

mora plays a pivotal role in formulating many linguistic phenomena in Japanese. In this sense, Japanese presents compelling evidence that the two prosodic units are not mutually exclusive in a single prosodic system.

5 Accent and intonation

5.1 Terminologies

Several chapters in this volume discuss word accent either as a main topic (Kawahara Ch. 11) or as a secondary topic in relation to the main subject of the chapter (Fujimoto Ch. 4; Nasu Ch. 6; Ito and Mester Ch. 7; Kubozono Ch. 8; Ito and Mester Ch. 9; Igarashi Ch. 13; Ishihara Ch. 14). This section summarizes some basic concepts and ideas that are commonly assumed in these chapters.

The word "accent" is often used ambiguously in the literature of Japanese phonetics and phonology. On the one hand, "accent" refers to the overall prosodic pattern, or accentuation, of the word: "accent" is used in this sense in such expressions as the 'accent of nouns', the 'accent of Tokyo Japanese', and the 'phonetics/phonology of Japanese accent'. On the other hand, the same word is also used to refer to the phonological prominence assigned to a particular position of the word, or "akusento-kaku" (accent kernel) in the more traditional terminologies (Hattori 1973; Uwano 1997, 2012). In Tokyo Japanese, this refers to an abrupt pitch drop in words like /i'.no.ti/ 'life' and /ko.ko'.ro/ 'heart'. Words with such a phonological/phonetic feature are called "accented" words as opposed to "unaccented" words like /ne.zu.mi/ 'mouse'.

These two meanings of the term are often used in one and the same article or book. One finds, for example, that words are classified into "accented" and "unaccented" words in an article entitled 'the accent of Tokyo Japanese'. A similar ambiguity can be found in the literature of English word accent, where the word "stress" is used in two ways: "Stress" in the 'stress system of English' refers to the overall prominence pattern of the word, whereas "stress" in the sentence 'the stress is on the initial syllable' refers to the phonological prominence given to a particular syllable of the word.

The term "accent" is used in the two senses in this volume, too, but we will try to disambiguate the two meanings as much as possible. We will, for example, use the terms "accent pattern" and "accentuation" if we mean the overall prominence pattern defined at the word level.

5.2 Accent system and representations

The word accent of Japanese is often called 'pitch accent' as opposed to 'stress accent' (Trubetzkoy 1958; Beckman 1986). This is based on the fact that word-level

phonological prominence is signaled primarily by pitch (F0, or fundamental frequency) rather than other phonetic parameters such as duration, intensity and vowel quality.

Tokyo Japanese is generally considered to have a multiple-pattern system ("takei-akusento", 多型アクセント) for nouns and a two-pattern system ("nikei-akusento", 二型アクセント) for verbs and adjectives (Uwano 1999). This means that the number of accent patterns increases in nouns as the word becomes phonologically longer, whereas verbs and adjectives have only two accent patterns no matter how long they may be. Specifically, nouns obey the (n+1) rule whereby monosyllabic, disyllabic and trisyllabic nouns have two, three, and four accent patterns, respectively. These are illustrated in (33)–(35), where /ga/ (nominative particle) is attached. For the sake of description, high-pitched moras are denoted by capital letters.

(33) a. hi'-ga (HI-ga) 'fire'
b. hi-ga (hi-GA) 'sun, sunshine'

(34) a. a'.me-ga (Ame-ga) 'rain'
ha'.na-ga (HA.na-ga) 'Hana, a girl's name'
b. ha.na'-ga (haNA-ga) 'flower'
c. a.me-ga (aME-GA) 'candy'
ha.na-ga (haNA-GA) 'nose'

(35) a. i'.no.ti-ga (I.no.ti-ga) 'life'
b. ko.ko'.ro-ga (ko.KO.ro-ga) 'heart'
c. o.to.ko'-ga (o.TO.KO-ga) 'man'
d. ne.zu.mi-ga (ne.ZU.MI-GA) 'mouse'

As shown in (33)–(35), the number of accent patterns is greater than the number of syllables by one. This is because of the unaccented pattern, or the pattern that does not involve a pitch drop even if a particle is attached to the noun.[12] In other words, nouns in Tokyo Japanese contrast not only in the position of the pitch accent but also in its presence or absence.

While an abrupt pitch drop is thus the primary acoustic correlate of the word accent in Tokyo Japanese, a pitch rise is a redundant feature. Namely, the position of pitch rise is fully predictable in this system: pitch rises in word-initial position unless the initial syllable is accented or, alternatively, the initial two moras must differ in pitch height (Haraguchi 1977). As noted by Kawakami (1961), Uwano (1977, 2012) and Pierrehumbert and Beckman (1988), however, the word-initial pitch rise is

12 Finally-accented and unaccented words cannot be distinguished from each other easily if pronounced in isolation, i.e. without the following particle (Uwano 1977; Vance 1995).

not a property of the word per se, but a property of the phrase. This can be seen in a phrase consisting of an unaccented adjective and a noun, where the word-initial pitch rise in the noun disappears. The phrase /ko.no ko.ko'.ro/ 'this heart', for example, does not show a pitch rise in the initial syllable of /ko.ko'.ro/ (unless this noun is emphasized for some reason): /koNO KOKOro/. This suggests that it is only the position of pitch fall that is marked or stored in the lexicon.

The relationship between the word and its accent pattern is basically arbitrary, so that it is difficult to predict which word takes which accent pattern. Naturally, there are many segmentally homophonous pairs of words that can be distinguished by word accent. Two points must be noted here. First, most minimal pairs of words that are distinguished by word accent involve a contrast in accentedness; that is, one word is accented and the other, unaccented. In contrast, minimal pairs of words that are distinguished by the position of accent are relatively few: e.g. /ha'na/ 'Hana' vs. /ha.na'/ 'flower'. This suggests that the presence or absence of word accent is primary, while its position is only secondary. This analysis is compatible with several facts to be mentioned shortly below.

Second, while there are many pairs of segmentally homophonous words that are distinguished by word accent, there are also many pairs that cannot be distinguished by word accent. These words are homophonous in the strict sense of the term. Some examples are given below.

(36) ku'.mo 'spider', 'cloud'
ko'o.sya 'the latter case', 'school building', 'public corporation', 'a tactful person'
su.ro'o 'throw', 'slow'

Surprisingly, this type of absolute homophones outnumbers homophones that can be distinguished by word accent: according to the statistical work by Sibata and Shibata (1990), only 14% of segmental homophones in Tokyo Japanese are distinguished by word accent, while the remaining 86% are completely homophonous. This suggests that the distinctive function is not the primary function of word accent in this dialect.

It is worth adding a few points about the (n+1) rule here. First, the permitted accent patterns do not occur equally frequently. On the contrary, only two patterns are very productive. In the case of three-mora nouns in (35), the pattern with an accent on the antepenultimate mora in (35a) and the unaccented pattern in (35d) are productive, while the other patterns – those with an accent on the penultimate or final mora – are much less productive. This can be seen clearly from the data in Table 6.

Table 6: Accent patterns and their frequencies in three-mora nouns (Kubozono 2006: 15)

Accent pattern	Antepenultimate accent	Penultimate accent	Final accent	Unacceted
Frequency	42%	4%	2%	52%

Second, the discrepancy between the productive and unproductive patterns becomes more prominent as the word becomes longer. In fact, the (n+1) rule is no longer valid for five-mora or longer nouns. A majority of long loanwords, for example, have an accent on their antepenultimate mora unless they are unaccented. This is basically true of native and SJ words, too. Most native and SJ words are morphologically complex, i.e. compound nouns, if they are longer than four moras, and are hence subject to compound accent rules. Nevertheless, the compound accent rules for nouns produce only two major patterns, those with a compound accent on the third or fourth mora from the end of the word, and those that are unaccented. Homophonous pairs show this.

(37) a.ki.ta'-ken 'Akita Prefecture' vs. a.ki.ta-ken 'Akita Dog'
si.ke'n-kan 'examiner' vs. si.ken-kan 'test tube'
i.wa.te'-san 'Mt. Iwate' vs. i.wa.te-san 'made in Iwate, produce of Iwate'

The fact that only two accent patterns are productive in Tokyo Japanese has led Kubozono (2008a,b) to propose that nouns in this dialect basically constitute a two-pattern system with a certain number of lexical exceptions, or words whose accent position is particularly marked in the lexicon. Under this analysis, nouns in this dialect contrast primarily in the presence or absence of word accent, and not in its position, just like verbs and adjectives described in (38). This is compatible with the idea mentioned above, that presence or absence is the primary feature of word accent, whereas position is only secondary.

(38) a. accented verbs and adjectives
na'.ru 'to be completed', ha.re'.ru 'to clear up', a.tu'-i 'hot', a.o'-i 'blue'

b. unaccented verbs and adjectives
na.ru 'to ring', ha.re.ru 'to become swollen', a.tu-i 'thick', a.ka-i 'red'

5.3 Accent rules

We have seen that loanwords and relatively long native and SJ words exhibit a limited number of accent patterns albeit the (n+1) rule. This suggests that the accentuation of Tokyo Japanese is rule-governed to a large extent. In the literature, in fact, one finds a number of accent rules, including the rule for loanwords, a set of accent

rules for compound nouns, and the accent rule for verbs and adjectives (Akinaga 1985, 2001). Endeavors have been made in the past few decades to 'refine' these rules (see Kawahara Ch. 11, this volume, for a full discussion).

Scholars have attempted to generalize the various accent rules. To take one example, the loanword accent rule formulated in (29) has been shown to account for the accentuation of most *accented* native and SJ words, too (Kubozono 2006). Native and SJ words differ from loanwords in the abundance of unaccented words, but if we focus on accented words, they do not essentially differ from loanwords. In other words, the loanword accent rule is not a rule specifically for loanwords, but is a general rule for accented nouns in Tokyo Japanese.

The accent rules for compound nouns have also been generalized to a considerable extent. It was long believed that compound nouns fall into two groups depending on their accentual behaviors: those with a short (monomoraic or bimoraic) second member and those with a long (trimoraic or longer) second member (Akinaga 1985; McCawley 1968; Poser 1990; Uwano 1997). In this analysis, the first group of compound nouns was further classified into three subgroups according to their accent patterns: (i) those that attract a compound accent on the final syllable of the first member, e.g. /a.ba.re'-u.ma/ 'spirited horse', (ii) those with a compound accent on the initial syllable of the second member, e.g. /pe.ru.sja-ne'.ko/ 'Persian cat', and (iii) those that are unaccented, e.g. /ne.zu.mi-i.ro/ 'rat, color; grey'. On the other hand, compound nouns with a long second member were classified into two groups: (iv) those with a compound accent on the initial syllable of the second member, e.g. /too.kjoo-da'i.ga.ku/ 'Tokyo University', and (v) those that keep the lexical accent of the second member, e.g. /ja.ma.to-na.de'.si.ko/ 'Japanese woman'. This means that five major accent patterns or rules were put forward for compound nouns in Tokyo Japanese.

Kubozono (1995b, 1997) challenged this traditional analysis and proposed that the seemingly different accent patterns can be generalized. The key to this new analysis is to differentiate rule-governed patterns from lexically-marked patterns. In the case under consideration, the pattern in (iii) must be removed from the scope of accent computation since the unaccented pattern is due to the deaccenting effect of a certain number of morphemes that must be specified as such in the lexicon. Once this lexically-marked pattern is removed, all other compound accent patterns can more or less be generalized in an output-oriented OT analysis – they can be accounted for as the result of interactions between several general constraints, notably 'nonfinality' (don't put an accent at the end of the word), 'edgemostness' (put an accent maximally towards the end of the word) and 'Max-accent' (preserve the original accent of the second member).

The accent patterns of compound nouns can be generalized not only with each other but with morphologically simplex (accented) nouns, too (Kubozono 2002, 2008a,b). Furthermore, they can also be generalized with the accent pattern of accented verbs and adjectives, and also with the famous accent rule of Latin

(Kubozono 2006, 2008a,b, 2011). The key to this new generalization is to separate the unaccented pattern clearly from accented patterns.

The efforts to generalize seemingly different accent patterns are closely linked to those to clarify the extent to which the patterns are predictable. In the case of trimoraic nouns in (35), (35a) and (35d) are much more productive than the other two patterns, as already mentioned. If we focus on the accented patterns, we can derive the accent pattern in (35a) by rule – actually by the same rule that governs loanwords – while attributing the other accented patterns to lexical exceptions. Again, the crucial point of this analysis is to differentiate rule-governed patterns from lexically-marked ones.

One might argue against this analysis that it is still impossible to explain why /i'noti/ 'life' in (34a) belongs to the rule-governed group, while /koko'ro/ 'heart' and /otoko'/ 'man' belong to the lexically-marked group. This is basically true, but it does not entail that the three patterns must be treated in the same way. To take one parallel example from the verb morphology of English, it is difficult to explain, at least in the synchronic grammar, why 'go' and 'run' take irregular verbal forms (*go-went-gone*, *run-ran-run*), while 'visit' and 'walk' are regular verbs (visit-visited-visited, walk-walked-walked). Yet, few people would argue against the standard analysis whereby the former verbs should be treated differently from the latter verbs: the past and perfect forms are marked in the lexicon for the former verbs but are derived by rule for the latter verbs. In the same way, while it is difficult to explain why /i'noti/ but not /koko'ro/ and /otoko'/ has an accent on the antepenultimate mora, it is nevertheless reasonable to assume that one of the accent patterns is derived by rule, while the other patterns are lexically marked.

It is worth mentioning here that serious attempts have been made to account for the distribution of unaccented words. In the case of native and SJ words, it is difficult to explain why words like /nezumi/ 'mouse' in (35d) take this accent pattern. However, analyses of loanwords and compound words have clarified the phonological conditions on this accented pattern (Kubozono, 1996, 2010; Kubozono, Ito, and Mester 1997; Giriko 2009). For example, it is now clear that loanwords tend to take the unaccented pattern if they are four moras long and involve a sequence of two light syllables in word-final position (see Table 3 above). If further developed, this approach to the unaccented pattern might be able to define the linguistic conditions under which this peculiar accent pattern occurs in the Japanese vocabulary as a whole, which has been a long-standing mystery in Japanese phonology.

Finally, the attempts to generalize various accent rules are mingled with the approaches to discovering the general principles underlying the accent rules. Take the antepenultimate rule in (29), for example. As mentioned above, this rule has been shown to account for the accentuation of most accented nouns in Tokyo Japanese. It has also been demonstrated that it is crucially similar to the famous accent rule of Latin (Kubozono 1996, 2008a,b, 2011). However, it is also assumed in more theoretical analyses (Katayama 1998; Shinohara 2000) that the antepenultimate

effect is an epiphenomenon that results from the interaction of several independent, general principles. These theoretical analyses provide us with the ground where we can discuss Japanese accent from cross-linguistic and typological perspectives.

5.4 Accent and intonation

Word accent remains largely intact in the sentence level in Tokyo Japanese. Thus, the distinction between accented and unaccented words remains intact when they are embedded in sentences. In this sense, word accent is basically independent of sentence-level prosody. Yet, it is closely related with sentence intonation in several ways. One of the most important phenomena in this area is so-called downstep (or catathesis) by which pitch range is lowered after an accented word/phrase. In other words, one and the same word is realized at a lower pitch level if it is preceded by a lexically accented phrase than if it is preceded by a lexically unaccented one (Poser 1984; Pierrehumbert and Beckman 1988; Kubozono 1988).

(39) a. u'mai nomi'mono 'tasty drink'
amai nomi'mono 'sweet drink'
b. u'mai ame 'tasty candy'
amai ame 'sweet candy'
c. na'oko-no eri'maki 'Naoko's muffler'
naomi-no eri'maki 'Naomi's muffler'

In the pair of sentences in (39a), the second phrase /nomi'mono/ is realized considerably lower if preceded by the accented phrase /u'mai/ than if preceded by the unaccented phrase /amai/. Similarly, the unaccented phrase /ame/ in (39b) has considerably lower pitch values if it follows the accented phrase than if it follows the unaccented one. Overall, the accentedness of the first phrase affects the pitch of the following phrases significantly. This and other issues pertaining to word accent and intonation will be discussed in detail by Igarashi (Ch. 13) and Ishihara (Ch. 14).

6 Word types and word structure

6.1 Lexical strata

To understand the phonetic and phonological structures of Japanese, it is vital to understand the organization of its lexicon, or the lexical strata. Specifically, it is important to know that different types of words have different structures and often

exhibit different phonological patterns/behaviors. Words in Japanese are generally classified into three groups: (i) native words, (ii) Sino-Japanese (SJ) words, and (iii) loanwords. The third group usually refers to loanwords that have been borrowed from English and other language in the past few centuries. About 84% of loanwords in modern Japanese come from English (Sibata 1994). Some linguists add mimetic words as a fourth group (Ito and Mester 1999, 2008; Nasu 1999): see Nasu (this volume) for a full analysis of this type of words.

The three types of words show many linguistic differences. In terms of orthography, SJ words and loanwords are generally written in Chinese characters and *katakana* letters, respectively. Some loanwords, especially acronyms, are written in English alphabets, too: e.g. ANA 'All-Nippon Airways'. Native words, in contrast, are written either in *hiragana* letters or in Chinese characters (or the combination of the two): e.g. 学ぶ or まなぶ for the verb /manabu/ 'to learn'.

In addition to the differences in orthography, the three types of words in Japanese exhibit different linguistic structures and patterns. As for vowels, for example, not many native words have a long vowel or diphthong. Many SJ words and loanwords have a long vowel or diphthong, but long vowels in SJ words are restricted to /oo/, /ee/ and /uu/, as noted above. Loanwords have all five long vowels since tense vowels in English are generally adapted as long vowels in Japanese (see (7) above).

As for consonants, /p/ occurs quite commonly in loanwords but not in native words. This is due to the historical changes whereby /p/ turned into other sounds – [h] in word-initial position and [w] in medial position – in the course of the history. In native and SJ words, in fact, /p/ is found only as a geminate consonant in word-medial positions. This is clearly shown by the morphophonemic alternation between /h/ in word-initial position and /p/ in medial position: e.g. /<u>hjo</u>o/ 'table, chart' vs. /ha<u>p.pjo</u>o/ 'presentation', /<u>hi</u>.ru.ma/ 'daytime' vs. /ma<u>p.pi</u>.ru.ma/ 'midday'.

6.2 Phonological length

The three types of words differ in phonological length, too. As for the size of morphemes, SJ morphemes are the shortest: they are one or two moras long by mora count and one or two syllables long by syllable count. These length restrictions reflect the fact that morphemes in Chinese are basically monosyllables, some of which which become disyllabic via vowel epenthesis (see (2) above). Not many SJ morphemes are used as independent words in Japanese.[13] Rather, two morphemes are usually combined to form a 'word' in the language: e.g. /gaku-mon/ 'learning, to ask; learning, science'.

13 Some SJ morphemes can be used independently: e.g. /ni.ku/ 'meat', /e.ki/ 'station', /ki.ku/ 'chrysanthemum', /hu.ku/ 'clothes', /hu.ku/ 'happiness'. Many of these morphemes tend to be intuitively taken as native morphemes by native speakers.

Native morphemes can be longer than SJ morphemes, but they do not usually exceed three moras. Seen conversely, most four-mora or longer native words are compounds, at least etymologically. Thus, the noun /mi.zu.u.mi/ 'lake' consists of two morphemes, /mizu/ 'water' and /umi/ 'sea'. The same is true of the four-mora verb, /nu.ka.zu.ku/ 'to prostrate oneself', where /nuka/ 'forehead' and /tuku/ 'to push, to prick' are combined. In comparison, loanwords can be quite long. They are minimally bimoraic, e.g. /ba.su/ 'bus' and /pin/ 'pin',[14] but can be many moras long. For example, /kon.pjuu.taa/ 'computer' is a monomorpheme in Japanese that consists of six moras.

It is probably worth adding here that four-mora length is the most popular length of words in the Japanese vocabulary. In Yokoyama's (1979) data, cited in Hayashi (1982), four-mora words account for 49% of all the words listed in Sanseido's dictionary (*Shinmeikai Kokugo Jiten*), followed by three-mora words (30%) and five-mora words (9%) (Table 7).

Table 7: Frequency of Japanese words as a function of their length (based on Yokoyama 1979)

No. of moras	1	2	3	4	5	6	7	8~
No. of words	282	3,785	16,095	26,559	4,859	1,858	471	278
%	0.5	7.0	29.7	49.0	9.0	3.4	0.9	0.5

Similar statistics were reported by Hayashi (1957), who looked at about 47,000 words listed in the 1951 version of the NHK Accent Dictionary.

Table 8: Frequency of Japanese words as a function of their length (Hayashi 1957)

No. of moras	1	2	3	4	5	6	7	8~
%	0.3	4.8	22.7	38.8	17.7	11.0	3.3	1.5

6.3 Accent

The three types of words often display different phonological patterns. One of the most remarkable differences in word accent can be found in the frequency of the unaccented accent pattern. This pattern is very popular in native and SJ words. According to Kubozono (2006), it accounts for about 71% and 51% of trimoraic native and SJ words, respectively. In contrast, the same accent pattern is relatively rare in loanwords: it accounts for only 10% of loanwords as a whole (Sibata 1994) and 7% of trimoraic loanwords (Kubozono 2006). The huge difference between native/SJ words and loanwords with respect to the popularity of the unaccented word has been a matter of debate in Japanese phonology. In loanword phonology,

14 Musical notes /do re mi/ may be the only exceptions.

in fact, it has been a mystery why loanwords behave so differently from the other two types of words and, more specifically, where the loanword accent rule in (29) comes from (Sibata 1994; Shinohara 2000; Kubozono 2006).

Apart from this, the accent patterns of nouns have much to do with their phonological length. For example, the unaccented pattern is most common in four-mora words in all three types of words. It remains a mystery why words of this phonological length favor the unaccented pattern (Ito and Mester, forthcoming).

6.4 Rendaku

The three types of words behave quite differently in the phonological process of *rendaku*, too. *Rendaku* is a process converting voiceless obstruents into voiced ones at the beginning of the second elements of compounds: see Vance (this volume) for a full discussion. This process is very productive in compound words whose second member is a native morpheme. In contrast, it occurs infrequently in SJ words and almost never in loanwords. Take the homophonous words, /kai/ 'shell' and /kai/ 'party', the first of which is a native morpheme and the second is a SJ morpheme. *Rendaku* readily applies to the first, as in (40a), but not to the second, as in (40b).

(40) a. sakura + kai → sakura-gai 'cherry shell'
nimai + kai → nimai-gai 'bivalve'
hora + kai → hora-gai 'trumpet shell'

b. owakare + kai → owakare-kai 'farewell party'
doosoo + kai → doosoo-kai 'class reunion'
kangei + kai → kangei-kai 'welcome party'

Similarly, *rendaku* fails to apply to loanwords, as can be seen from the comparison of /ka.me.ra/ 'camera' (loanword) and /ka.me/ 'turtle' (native word).

(41) a. de.zi.ta.ru + ka.me.ra → de.zi.ta.ru-ka.me.ra 'digital camera'
tu.kai.su.te + ka.me.ra → tu.kai.su.te-ka.me.ra 'disposable camera'

b. u.mi + ka.me → u.mi-ga.me 'sea turtle'
zoo + ka.me → zoo-ga.me 'elephant tortoise'

7 Broader perspectives

In the foregoing discussion, we have described the basic phonetic and phonological structures of modern Tokyo Japanese and defined basic concepts and notions that are often used in this volume. In passing, we have also referred to the chapters in

Parts II–IV. In the last part (Part V) of the volume, we have four chapters that look at Japanese phonetics and phonology from broader perspectives, in the interface with other subfields of linguistics such as historical and corpus linguistics, L1 phonology acquisition, and L2 research. These four chapters are briefly summarized below.

7.1 Historical phonology

Historical studies of the Japanese language provide rich resources that are significant and useful for understanding the structure of modern Japanese. The Japanese language is attested back to the eighth century with a large amount of written records and dialectal information, although historical linguistic research faces many challenges. The extensive literature on Japanese historical phonology is not well known to linguists except for specialists in the history of Japanese. This chapter (Chapter 15) reviews the results and the points of controversy that have emerged from studies in the last few decades with main focus on the question of what historical studies reveal about the structure of modern Japanese.

7.2 Corpus-based phonetics

Chapter 16 examines how corpus-based quantitative studies can contribute to the better understanding of the phonetic/phonological structure of modern Japanese. After a retrospective survey of the history of speech corpus development and corpus-based analyses in the fields of phonetics and speech processing, we will describe the Corpus of Spontaneous Japanese (CSJ), which is the most representative speech corpus of the present-day spoken Japanese with respect to the corpus size, richness of annotation, and the degree of dissemination to the research communities. We will then consider several cases showing how CSJ can be applied to the quantitative study of speech processing and phonetics of spontaneous speech. We will conclude the chapter with some prospective discussions of the coming development of corpus-based phonetics in Japan and the world.

7.3 L1 Phonology: phonological development

Chapter 17 has two purposes. First, it presents an overview of descriptive findings on the phonological acquisition of Japanese as a native language. Drawing on classic as well as more recent developmental work, the chapter will exemplify how children exposed to Japanese acquire various aspects of its phonology, including segmental contrasts, durational contrasts, alternations, morphophonemic processes, word segmentation and pitch phonology. Second, it discusses the implications of these findings

for our understanding of the phonological structure of Japanese and general phonological theory as well as models of phonological development. The chapter concludes with suggestions for future directions of investigation.

7.4 L2 Phonetics and phonology

Chapter 18 introduces research in phonetic and phonological aspects of Japanese as learned by non-native speakers of Japanese, and contextualizes it in a broader field of second language (L2) speech acquisition. The first two sections introduce extant research in elements of Japanese speech sounds that are difficult for learners to produce and perceive, e.g., phonemic length contrasts of vowels and consonants, pitch accent, and other segmental contrasts. The next section introduces extant theories of L2 speech acquisition, in which various factors predict degrees of success in L2 acquisition, and reviews studies exploring different types of training and multimodal learning methods to enable learners to acquire difficult L2 speech sounds. The final section discusses areas of L2 Japanese that need investigations in the future.

8 Phonetic symbols

8.1 IPA symbols

Before concluding this introductory chapter, we give a guide to the phonetic and phonemic symbols that are used in this volume. As for phonetic representations, we broadly follow IPA for phonetic symbols. The following shows the phonemic and phonetic representations for sounds that are often symbolized in different ways in different books/articles. These representations are used in all chapters of this volume unless otherwise stated in the chapter.

(42)	/u/	[ɯ]
	/si/	[ʃi]
	/ti/	[tʃi]
	/tu/	[tsɯ]
	/hi/	[çi]
	/hu/	[ɸɯ]
	/ra/	[ra]
	/ja/	[ja]

There are some sound sequences that are observed only in loanwords. These are represented with the IPA symbols closest to them: e.g. [ti] in [pa:ti:] ‘party’.

8.2 Long vowels and coda consonants

Long vowels are represented in various ways in the literature. In this volume, we chose to double the same symbol for phonemic representations, e.g. /kookoo/ 'high school', avoiding the archiphonemic representation (/koR/ or /koH/) as much as possible. Long vowels are shown with a length marker in phonetic representations: e.g. [ko:ko:] 'high school'.

Nasals and obstruents in the coda position are often called moraic nasals and obstruents, respectively, in Japanese phonology. Coda nasals are shown with the symbol /n/ instead of the archiphonemic /N/ in phonemic representation. In phonetic accounts, they are represented as [m], [n] or [ŋ], so that they reflect the actual place of articulation involved. Coda obstruents are written with double consonants in both phonemic and phonetic representations: e.g. /kitte/ [kitte] 'postal stamp'. Here, too, the archiphonemic symbol /Q/ (e.g. /kiQ.te/ 'postal stamp') is avoided as much as possible.

8.3 Accent and syllable boundaries

Word accent has been represented in many ways in the literature. What is relevant in the phonological description of word accent in Tokyo Japanese is the position of pitch accent (or accent kernel, "akusento-kaku"), or the position where an abrupt pitch fall occurs. This position is marked by an apostrophe (') rather than a diacritic like (*) or an accent mark on the vowel, e.g. /í/. Thus, /tookjoo-daigaku/ 'Tokyo University' is represented as /tookjoo-da'igaku/.

Unaccented words are often shown with no accent mark, but /^{0}/ is added to the end of the word if it is necessary to show its lack of accent explicitly: e.g. /daigaku^0/ 'university'.

Dots (.) indicate syllable boundaries. This symbol is used wherever necessary: e.g. /dai.ga.ku/ 'university'. Hyphens (-) are often used to denote mora boundaries in the discussion of moraic structure: e.g. /da-i-ga-ku/. They are also used to show morpheme boundaries in compounds.

8.4 Other terminologies

If more than one terminology or expression is used in the literature, one of them is consistently used here to give uniformity to the whole volume. This includes the following terms/expressions: those in the parentheses are not used in this volume.

(43) moras (vs. morae)
downstep (vs. catathesis)
major phrase (vs. intermediate phrase)
minor phrase (vs. accentual phrase)

8.5 Romanization

Romanization of Japanese words can be a source of confusion. In order to achieve a consistency, this volume – and the entire handbook series – follows the style sheet of *Gengo Kenkyu*, the journal of the Linguistic Society of Japan. Its basic rule is to use *kunrei-shiki* for linguistic samples and *Hebon-shiki* for references. Details can be found at the following website (Japanese and English, respectively):
http://www3.nacos.com/lsj/modules/documents/LSJpapers/j-gkstyle2010.pdf
http://www3.nacos.com/lsj/modules/documents/LSJpapers/j-gkstyle2010e.pdf

Acknowledgements

We would like to thank the following people for their invaluable help in editing this volume: Donna Erickson, Hyun-Kyung Hwang, Clemens Poppe, Yasunori Takahashi, and Izumi Takiguchi. This volume was financially supported by the NINJAL collaborative research project 'Phonological characteristics of the Japanese lexicon' and the JSPS grant-in-Aid for Scientific Research (Grant no. 22242011).

References

Akinaga, Kazue. 1985. Kyōtsūgo no akusento [Accent of standard Japanese]. In NHK (ed.), *Pronunciation and accent dictionary* [The Japanese language pronunciation and accent dictionary]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Akinaga, Kazue. 2001. Tōkyō akusento no shūtoku hōsoku [Rules for learning Tokyo accent]. In Haruhiko Kindaichi (ed.), *Shinmeikai nihongo akusento jiten* [Shinmeikai Japanese accent dictionary], 10–106. Tokyo: Sanseido.

Árnason, Kristján. 1980. *Quantity in historical linguistics*. Cambridge: Cambridge University Press.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Bekku, Sadanori. 1977. Nihongo no rizumu [Rhythm in Japanese]. Tokyo: Kodansha.

Bloch, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26. 86–125.

Campbell, Nick. 1992. Segmental elasticity and timing in Japanese. In Yoh'ichi Tohkura, Eric V. Vatikiotis-Bateson and Yoshinori Sagisaka (eds.), *Speech perception, production and linguistics structure*, 403–418. Tokyo: Ohmsha.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Giriko, Mikio. 2009. *Nihongo ni okeru go ninshiki to heiban-gata akusento* [Word structure recognition and the unaccented pattern in Japanese]. Kobe: Kobe University dissertation.

Han, Mieko Shimizu. 1962. *Japanese phonology: An analysis based upon sound spectrograms.* Tokyo: Kenkyusha.

Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology.* Tokyo: Kaitakusha.

Hattori, Shiro. 1973. Akusento towa nani ka. Sono benbetsuteki tokuchō towa [What is the prosodeme? And what is its distinctive feature?]. *Gengo no Kagaku* 4. 46–53.

Hayashi, Ōki. 1957. Goi [Vocabulary]. In Etsutarō Iwabuchi, Ōki Hayashi, Hatsutarō Ōishi and Takesi Sibata (eds.), *Kōza gendai kokugogaku II: Kotoba no taikei* [Series in modern Japanese linguistics 2: The system of language], 95–125. Tokyo: Chikuma Shobō.

Hayashi, Ōki (ed.). 1982. *Zusetsu nihongo* [An illustrated guide to Japanese]. Tokyo: Kadokawa.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 53–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: Chicago University Press.

Hirata, Yukari. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32. 565–589.

Hirata, Yukari. this volume. Chapter 18: L2 phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Hirata, Yukari and Kimiko Tsukada. 2009. Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* 66(3). 129–149.

Homma, Yayoi. 1984. The rhythm of tanka – Short Japanese poems. *Proceedings of the 13th International Congress of Linguists*, Tokyo, 618–624.

Igarashi, Yosuke. this volume. Chapter 13: Intonation. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ishihara, Shinichiro. this volume. Chapter 14: Syntax-phonology interface. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko. 1990. Prosodic minimality in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*. 213–239.

Ito, Junko and Armin Mester. 1993. Licensed segments and safe paths. *Canadian Journal of Linguistics* 38. 197–213.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *Phonology handbook*, 817–838. Oxford: Blackwell.

Ito, Junko and Armin Mester. 1999. The Structure of the phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Malden, MA and Oxford: Blackwell.

Ito, Junko and Armin Mester. 2003. Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.), *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*, 26–65. Tokyo: Kaitakusha.

Ito, Junko and Armin Mester. 2008. Lexical classes in phonology. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 84–106. Oxford: Oxford University Press.

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure* (MIT Press Linguistic Inquiry Monograph Series 41). Cambridge, MA: MIT Press.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko and Armin Mester. Forthcoming. Unaccentedness in Japanese. Submitted to *Linguistic Inquiry*.

Ito, Junko, Yoshihisa Kitagawa and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology*. Santa Cruz, CA: University of California Santa Cruz dissertation.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kawagoe, Itsue and Masako Arai. 2002. Shakuyōgo ni okeru sokuon [Consonant gemination in loanwords]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 6(1). 53–66.
Kawagoe, Itsue and Akiko Takemura. 2013. Geminate judgment of English-like words by Japanese native speakers: Differences in the borrowed forms of "stuff" and "tough". *Journal of East Asian Linguistics* 22(4). 307–337.
Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kawakami, Shin. 1961. On the relationship between word-tone and phrase-tone in Japanese. *The Study of Sounds* (The Phonetic Society of Japan) 9. 169–177. Reprinted in Shin Kawakami 1995.
Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to Japanese phonetics]. Tokyo: Ōfūsha.
Kawakami, Shin. 1995. *Nihongo akusento ronshū* [Papers on Japanese accent]. Tokyo: Kyūko Shoin.
Kindaichi, Haruhiko. 1967. *Nihongo on'in no kenkyū* [Studies in Japanese phonology]. Tokyo: Tōkyōdō Shuppan.
Kinoshita, Keisuke, Dawn Behne and Takayuki Arai. 2002. Duration and F0 as perceptual cues to Japanese vowel quantity. *Proceedings of the International Conference on Spoken Language Processing* (ICSLP), Denver, 757–760.
Komatsu, Hideo. 1981. *Nihongo no sekai 7: Nihongo no on'in* [The world of Japanese 7: Japanese phonology]. Tokyo: Chūōkōronsha.
Kubozono, Haruo. 1985. Speech errors and syllable structure. *Linguistics and Philology* 6. 220–243.
Kubozono, Haruo. 1988. *The organization of Japanese prosody*. Edinburgh: University of Edinburgh dissertation. (Kurosio, 1993).
Kubozono, Haruo. 1989. The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech* 32(3). 249–278.
Kubozono, Haruo. 1995a. *Gokeisei to on'in kōzō* [Word formation and phonological structure]. Tokyo: Kurosio.
Kubozono, Haruo. 1995b. Constraint interaction in Japanese phonology: Evidence from compound accent. In Rachel Walker, Ove Lorentz and Haruo Kubozono (eds.), *Phonology at Santa Cruz* 4, 21–38. Linguistics Research Center, UC Santa Cruz.
Kubozono, Haruo. 1996. Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin* (The Phonetic Society of Japan) 211. 71–82.
Kubozono, Haruo. 1997. Lexical markedness and variation: A non-derivational account. *Proceedings of the West Coast Conference on Formal Linguistics 15*, 273–287.
Kubozono, Haruo. 1999a. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Malden, MA and Oxford: Blackwell.
Kubozono, Haruo. 1999b. *Nihongo no onsei* [The sounds of Japanese]. Tokyo: Iwanami.
Kubozono, Haruo. 2000. Nihongo no goi to onsetsu kōzō [Japanese vocabulary and syllable structure]. *Nihongo Kenkyū* 20. 1–18.
Kubozono, Haruo. 2002. Syllable weight and Japanese accent. *Proceedings of Linguistics and Phonetics 2002*, Charles University Press (printed version) and Meikai University (CD-ROM version).
Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In Caroline Fery and Ruben van der Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.

Kubozono, Haruo. 2004. What does Kagoshima Japanese tell us about Japanese syllables? In Taro Kageyama and Hideki Kishimoto (eds.), *Nihongo no bunseki to gengo ruikei* [Analysis of Japanese and language typology], 75–92. Tokyo: Kurosio.

Kubozono, Haruo. 2005. [ai]-[au] asymmetry in English and Japanese. *English Linguistics* 22(1). 1–22.

Kubozono, Haruo. 2006. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.

Kubozono, Haruo. 2008a. *Akusento no hōsoku* [Laws of accent]. Tokyo: Iwanami.

Kubozono, Haruo. 2008b. Japanese accent. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 165–191. Oxford: Oxford University Press.

Kubozono, Haruo. 2008c. [ai]-[au] asymmetry: A phonetic account. In Haruo Kubozono (ed.), *Asymmetries in phonology: An east-Asian perspective*, 147–163. Tokyo: Kurosio.

Kubozono, Haruo. 2010. Accentuation of alphabetic acronyms in varieties of Japanese. *Lingua* 120. 2323–2335.

Kubozono, Haruo. 2011. Japanese pitch accent. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 5*, 2879–2907. Malden, MA and Oxford: Wiley-Blackwell.

Kubozono, Haruo. 2012a. Word-level vs. sentence-level prosody in Koshikijima Japanese. *The Linguistic Review* 29. 109–130.

Kubozono, Haruo. 2012b. Varieties of pitch accent systems in Japanese. *Lingua* 122. 1395–1414.

Kubozono, Haruo. 2013. Japanese word accent. In Mark Aronoff (ed.), *Oxford bibliographies in linguistics*. New York: Oxford University Press.

Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kubozono, Haruo, Junko Ito and Armin Mester. 1997. On'inkōzō kara mita go to ku no kyōkai: Fukugō meishi akusento no bunseki [The phonological boundary between the word and the phrase: An analysis of Japanese compound noun accentuation]. In Spoken Language Working Group (ed.), *Bunpō to onsei* [*Speech and grammar*], 147–166. Tokyo: Kurosio.

Kubozono, Haruo, Hajime Takeyasu and Mikio Giriko. 2013. On the positional asymmetry of consonant gemination in Japanese loanwords. *Journal of East Asian Linguistics* 22(3), 339–371.

Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. Cambridge, MA: MIT dissertation.

Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Lovins, Julia B. 1975. *Loanwords and the phonological structure of Japanese*. Bloomington: Indiana University Linguistics Club.

Macken, Marlys A. 1980. Aspects of the acquisition of stop systems: A cross-linguistic perspective. In Grace H. Yeni-Komshian, James F. Kavanagh and Charles A. Ferguson (eds.), *Child phonology (Volume 1 Production)*, 143–168. New York: Academic Press.

Maddieson, Ian. 1984. *Patterns of sounds*. New York: Cambridge University Press.

Maekawa, Kikuo. this volume. Chapter 16: Corpus-based phonetics. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Martin, Samuel E. 1975. *A reference grammar of Japanese*. New Haven: Yale University Press.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCawley, James D. 1978. What is a tone language? In Victoria A. Fromkin (ed.), *Tone: A linguistic survey*, 113–131. New York: Academic Press.

Mori, Yoko. 2002. Lengthening of Japanese monomoraic nouns. *Journal of Phonetics* 30(4). 689–708.

Nasu, Akio. 1999. Onomatope ni okeru yūseika to [p] no yūhyōsei [Voicing in onomatopoeia and the markedness of [p]]. *Onsei Kenkyū* [The Journal of the Phonetic Society of Japan] 3(3). 52–66.
Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Ohno, Kazutoshi. 2000. The lexical nature of rendaku in Japanese. *Japanese Korean Linguistics* 9. 151–164.
Onishi, Masao. 1937. Go'on hindo yori mitaru jikkakokugo no hatsuon kitei [The basis of pronunciation in ten languages seen from the frequencies of speech sounds]. *Onsei no Kenkyū* [The Study of Sounds, The Phonetic Society of Japan] 6. 153–181.
Ota, Mitsuhiko. this volume. Chapter 17: L1 phonology: phonological development. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Otake, Takashi, Giyoo Hatano, Anne Cutler and Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32. 258–278.
Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Port, Robert. F., Jonathan Dalby and Michael O'Dell. 1987. Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America* 81(5). 1574–1585.
Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.
Poser, Willliam J. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105.
Rosen, Eric. 2003. Systematic irregularity in Japanese *rendaku*: How the grammar mediates patterned lexical exceptions. *Canadian Journal of Linguistics* 48. 1–37.
Saito, Yoshio. 1997. *Nihongo onseigaku nyūmon* [Introduction to Japanese phonetics]. Tokyo: Sanseido.
Sherer, Tim D. 1994. *Prosoic phonotactics*. Amherst, MA: University of Massachusetts dissertation.
Shinohara, Shigeko. 2000. Default accentuation and foot structure in Japanese: Evidence from Japanese adaptations of French words. *Journal of East Asian Linguistics* 9. 55–96.
Sibata, Takesi. 1962. On'in [phonology]. In Kokugo gakkai (ed.), *Hōgengaku gaisetsu [Introduction to dialectoloy]*. Tokyo: Musashino Shobō.
Sibata, Takesi. 1994. Gairaigo ni okeru akusento kaku no ichi [The location of accent in loanwords]. In Kiyoji Sato (ed.), *Kokugo ronkyu 4: Gendaigo hōgen no kenkyū* [Studies in Japanese 4: Studies in modern Japanese and dialects], 338–418. Tokyo: Meiji Shoin.
Sibata, Takesi and Ritei Shibata. 1990. Akusento wa dōongo o dono teido benbetsu shiuru ka [To what extent can word accent distinguish homophones?]. *Keiryō Kokugogaku* 17(7). 317–327.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Tateishi, Koichi. 1989. Theoretical implications of the Japanese musician's language. *Proccedings of WCCFL* 8, 384–398. Stanford: Stanford Linguistic Association.
Trubetzkoy, Nikolai S. 1958/1969. *Grundzüge der Phonologie [Principles of phonology]*. Los Angeles: University of California Press.
Uwano, Zendo. 1977. Nihongo no akusento [Accent in Japanese]. In Susumu Ōno and Takesi Sibata (eds.), *Iwanami kōza nihongo 5: On'in* [Iwanami Japanese series 5: Phonology], 281–321. Tokyo: Iwanami.
Uwano, Zendo. 1997. Fukugō meishi kara mita nihongo shohōgen no akusento [Accent of various dialects of Japanese seen from the viewpoint of compound nouns]. In Tetsuya Kunihiro, Hajime Hirose and Morio Kohno (eds.), *Accent, intonation, rhythm and pause*, 231–270. Tokyo: Sanseido.

Uwano, Zendo. 1999. Classification of Japanese accent systems. In Shigeki Kaji (ed.), *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics'*, 151–186. Tokyo: ILCAA.

Uwano, Zendo. 2012. Three types of accent kernels in Japanese. *Lingua* 122. 1415–1440.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. New York: State University of New York Press.

Vance, Timothy J. 1995. Final-accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics* 23. 487–499.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.

Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Warner, Natasha and Takayuki Arai. 2001. Japanese mora-timing: A review. *Phonetica* 58. 1–25.

Yavaş, Mehmet. 1998. *Phonology: Development and disorders*. San Diego and London: Singular Publishing.

Yokoyama, Shoichi. 1979. Jisho midashigo no akusento no seikaku ni tsuite [On the nature of the accent of dictionary-listed words]. Conference handbook, The Acoustical Society of Japan.

Zec, Draga. 1995. The role of moraic structure in the distribution of segments within syllables. In Jacques Durand and Francis Katamba (eds.), *Frontiers of phonology: Atoms, structure, derivation*, 149–179. New York: Longman.

II Segmental phonetics and phonology

Shigeto Kawahara

1 The phonetics of *sokuon*, or geminate obstruents

1 Introduction

Japanese has a phonemic contrast between short and long nasal and obstruent consonant series, as exemplified by minimal pairs like [ka**t**a] 'frame' vs. [ka**tt**a] 'bought' and [ha**t**o] 'dove' vs. [ha**tt**o] 'hat'.[1] Short consonants are generally called "singletons", whereas long consonants are called "geminates", geminate obstruents, or obstruent geminates (see also Kawagoe, this volume). In the traditional literature on Japanese phonetics and phonology, the first half of obstruent geminates is called "*sokuon*" for which the symbol /Q/ is often used; in the Japanese orthographic system, this coda part is represented by "small tsu". Nasal geminates or their coda portions are called *hatsuon*; in the traditional literature they are represented by /N/. This chapter focuses on obstruent geminates. Henceforth, the term "geminate" refers specifically to obstruent geminates or *sokuon*, unless otherwise noted. This chapter provides an overview of the acoustic, perceptual, and articulatory characteristics of Japanese geminates.[2]

1 There is no phonemic contrast between short and long approximants (liquids and glides) in Japanese (see Kawagoe, this volume). Geminate approximants can however occur in emphatic forms (e.g. [kowwai] 'very scary' is an emphatic form of [kowai] 'scary'). See Aizawa (1985), Kawahara (2001), Kawahara and Braver (2014), and section 5.3 for the non-structure preserving nature of this emphatic gemination in Japanese. For the phonetic reasons that may possibly underlie the prohibition against lexical approximant geminates, see Kawahara, Pangilinan, and Garvey (2011), Kawahara (2012), Podesva (2000) and Solé (2002).

2 Primarily due to limitation of the author's expertise, L2 learning of Japanese geminates is not covered in this paper. Readers are directed to the following references: Han (1992); Motohashi-Saigo and Hardison (2009); Oba, Brown, and Handke (2009); Tajima et al. (2008), several papers in a special issue of *Onsei Kenkyū* 11:1 (Kubozono 2007), those cited therein, as well as Hirata (this volume). Another topic that this chapter does not cover is a gemination pattern found in the process of loanword adaptation (e.g. [bakkɯ] 'back' < English *back*), which arguably has a perceptual basis (e.g. Kawagoe and Takemura 2013; Takagi and Mann 1994, though cf. Kubozono, Ito and Mester 2008). See Kawagoe (this volume) and Kubozono (this volume) for further discussion on this phenomenon.

This chapter does not deal with long vowels, although many issues discussed for geminate consonants in this paper are also relevant to long vowels. Here I list some key references. For general durational properties of long vowels in Japanese, see Braver and Kawahara (2014); Han (1962); Hoequist (1982); Kawahara and Braver (2013); Mori (2002) and Port, Dalby, and O'Dell (1987); for the effect of speech rate on long vowel production and perception, see Hirata (2004) and Hirata and Lambacher (2004); for secondary, non-durational acoustic correlates and their perceptual impacts, see Behne et al. (1999); Hirata and Tsukada (2009) and Kinoshita, Behne, and Arai (2002).

The structure of this paper is as follows. Section 2 discusses acoustic correlates of a singleton/geminate contrast in Japanese. The primary acoustic correlate exploited by Japanese speakers is constriction duration; other acoustic correlates include various durational correlates (e.g. duration of preceding vowel) and non-durational correlates (e.g. spectral properties in surrounding vowels). Section 2 also discusses other topics including the search for invariance and manner effects, as well as comparison of Japanese with other languages. Section 3 provides an overview of the experiments on the perception of geminates in Japanese. It discusses the effect of constriction duration as the primary perceptual cue, and also discusses how the duration of surrounding intervals affects the perception of geminates. Section 4 provides an overview of the literature on the articulation of Japanese geminates. Several issues that require further investigation are identified throughout the paper, and Section 5 raises several other issues that are not covered in the rest of the paper.

2 The acoustic characteristics of geminates in Japanese

2.1 The primary acoustic correlate: constriction duration

Japanese is often assumed to be a mora-timed language (see Warner and Arai 1999 for a review; see also Otake, this volume, on mora-timing); geminates are moraic, while singletons are not; for example, disyllabic words containing a geminate like [katta] 'bought' or [hatto] 'hat' have three moras. Reflecting their moraic nature, geminate consonants in Japanese have a longer consonantal constriction. Acoustically, the primary correlate of a singleton-geminate contrast is a difference in constriction duration – i.e. for stops, it is closure duration and for fricatives, it is frication duration. (In this paper, "duration" refers phonetic measures and "length" refers phonological contrast; "constriction" refers to both stop closure and narrow aperture for fricatives).[3]

Before proceeding to the discussion, there is one remark about what is meant by a particular acoustic correlate being "primary". The concept of being "primary" can mean several different things. A primary acoustic correlate can be used to mean an acoustic parameter that is invariant across speakers, speech styles, phonological contexts, or even across languages; a "primary" cue is also used to mean that it constitutes the most important perceptual cue for listeners, one that dominates other secondary cues (Lahiri and Hankamer 1988) so that secondary cues are only exploited when the target stimuli are ambiguous in terms of the primary cue, distributing around a range that is not found in natural speech (Hankamer, Lahiri, and

3 For affricates, the primary acoustic correlate seems to lie in the difference in the closure duration, and not in frication duration (Oba, Brown, and Handke 2009). See section 2.3.2.

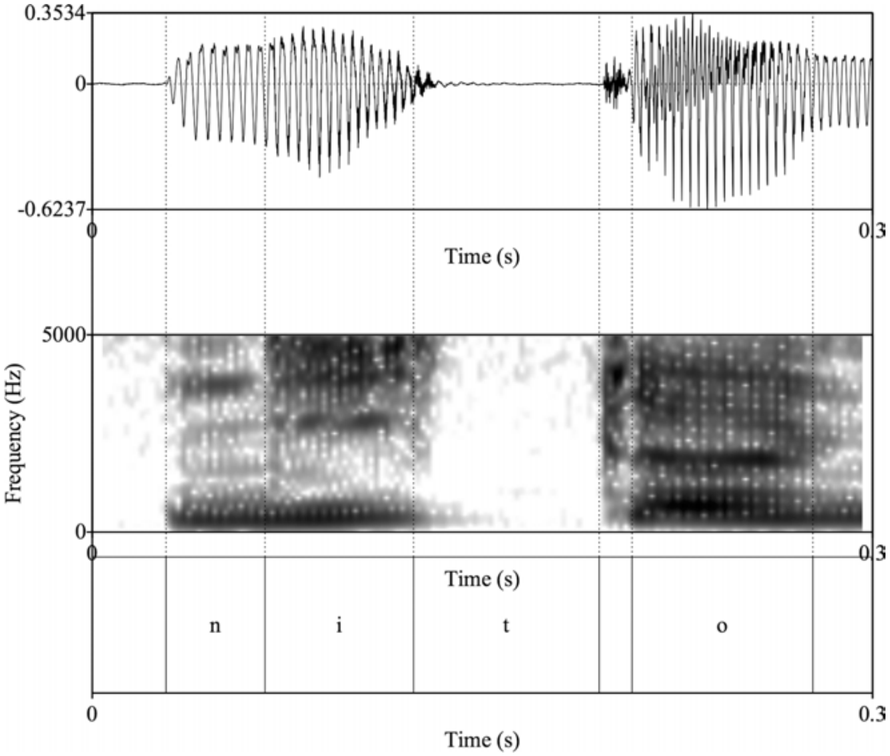


Figure 1: A singleton [t] in Japanese. Produced by a female native speaker of Japanese. The time scale is 300ms

Koreman 1989; Picket, Blumstein, and Burton 1999). For a general discussion on primacy of cues, see Abramson and Lisker (1985); Stevens and Blumstein (1981); Stevens and Keyser (1989); Whalen et al. (1993) and others; for a discussion of primacy in the context of length distinctions, see Abramson (1992); Hankamer, Lahiri, and Koreman (1989); Idemaru and Guion (2008); Lahiri and Hankamer (1988); Picket, Blumstein, and Burton (1999) and Ridouane (2010). Ridouane (2010) argues that cross-linguistically, differences in constriction duration are the most consistent acoustic correlates of singleton-geminate contrasts.

With this said, the primary acoustic correlate of Japanese geminates is greater duration compared to singletons: geminate consonants are characteristically longer than singleton consonants. Figures 1 and 2 show illustrative waveforms and spectrograms of a singleton [t] and a geminate [tt] in Japanese (with the same time scale of 300ms). As we can see, the geminate [tt] has a longer closure than the singleton [t].

Many acoustic studies have investigated the durational properties of singleton-geminate contrasts in Japanese, and Table 1 summarizes their findings. This summary shows that geminate stops are generally at least twice as long as corresponding singleton stops, and can sometimes be as three times as long, regardless of the

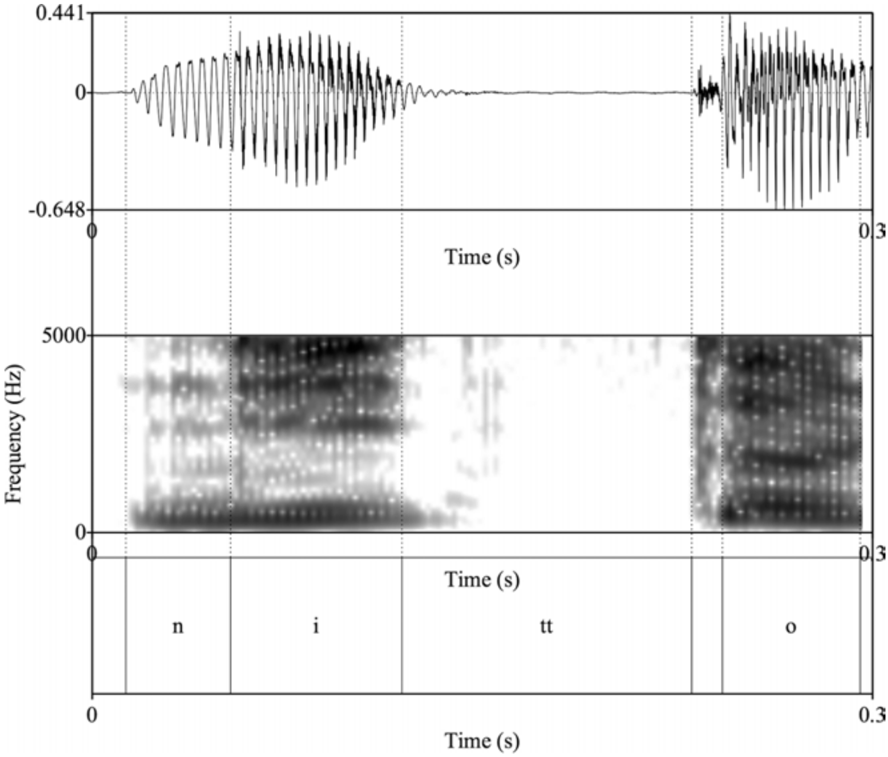


Figure 2: A geminate [tt]. The time scale is 300ms

place of articulation or voicing status of the consonants (though see section 2.3 for further discussion on the manner effect on geminate duration).

2.2 Secondary acoustic correlates

As with many other phonological contrasts, a singleton-geminate contrast is acoustically manifested not only by constriction duration, but by multiple other acoustic properties as well. (Multiplicity of acoustic correlates for phonological contrasts has been an important topic throughout the history of the phonetic theory; see, for example, Abramson 1998; Kingston and Diehl 1994; Lisker 1986 and references cited therein.)

2.2.1 Other durational correlates

In Japanese, vowels are longer before geminates than before singletons (Campbell 1999; Fukui 1978; Han 1994; Hirata 2007; Hirose and Ashby 2007; Idemaru and Guion

Table 1: Summary of the previous studies on closure duration of singleton and geminate stops and their ratios in Japanese. Duration measures are in miliseconds. SD = standard deviation; MoE = margin of error for 95% confidence intervals. Sing = singleton; Gem = geminate; VOT = Voice Onset Time; vls = voiceless; vcd = voiced

Sources	Sing duration	Gem duration	Ratio	Note
Han (1962)	–	–	2.6–3.0	based on small N
Homma (1981)	[p]: 77	[pp]: 183	2.38	4 speakers
	[b]: 55	[bb]: 159	2.89	
	[t]: 62	[tt]: 170	2.74	
	[d]: 35	[dd]: 144	4.11	
	[k]: 61	[kk]: 175	2.87	
	[g]: 41	[gg]: 134	3.27	
Beckman (1982)				(SD), 5 speakers
	[k]: 89 (17)	[kk]: 195 (32)	2.25	VOT included
	[k]: 64 (15)	[kk]: 171 (32)	2.79	VOT excluded
Port et al. (1987)				(SD), 10 speakers
	[k]: 65 (12)	[kk]: 149 (25)	2.29	ɯ_ ɯ
	[k]: 66 (14)	[kk]: 146 (28)	2.21	a_ ɯ
Han (1994)				(SD), 10 speakers
(see also Han 1992)	[p]: 76.3 (5.6)	[pp]: 195.9 (21.9)	2.57	sɯ_ai
	[p]: 72.9 (9.7)	[pp]: 205.4 (29.9)	2.82	sɯ_ori
	[t]: 71.5 (7.4)	[tt]: 192.3 (27.2)	2.69	i_e
	[t]: 53.5 (8.0)	[tt]: 166.6 (24.1)	3.11	ki_e
	[t]: 57.9 (10.2)	[tt]: 174.5 (21.5)	3.01	ʃi_ei
	[t]: 52.7 (8.0)	[tt]: 170.9 (25.8)	3.24	ki_e
	[t]: 68.2 (9.0)	[tt]: 189.8 (28.5)	2.78	i_a
	[k]: 63.5 (8.5)	[tt]: 178.2 (22.5)	2.81	yo_a
	[k]: 57.5 (8.5)	[tt]: 175.8 (30.9)	3.06	ʃi_e
	[k]: 79.4 (6.6)	[kk]: 198.7 (24.6)	2.50	ha_eN
Kawahara (2006a)	vls: 59.9 (2.1)	vls: 128.6 (3.1)	2.15	(MoE), 3 speakers
	vcd: 42.3 (1.7)	vcd: 113.1 (3.0)	2.67	
Hirose and Ashby (2007)	vls: 60.5	vls: 114.2	1.89	3 speakers
	vcd: 44	vcd: 108	2.45	
Idemaru & Guion (2008)	69 (28)	206 (45)	2.99	(SD), 6 speakers all stop consonants

2008; Kawahara 2006a, 2013b; Kawahara and Braver 2014; Ofuka 2003; Port, Dalby, and O'Dell 1987; Takeyasu 2012).[4] Port, Dalby, and O'Dell (1987) found, for example, that [ɯ] is on average 68ms before singleton [k] and 86ms before geminate [kk]; i.e. that [ɯ] is 18ms longer on average before geminates. Kawahara (2006a) found

4 Vowels are also longer in closed syllables before a so-called moraic nasal (or *hatsuon*) – i.e. in (C)VN – than in open syllables – i.e. in (C)V (Campbell 1999). This observation indicates that this lengthening is due to a general, syllable-based phenomenon. The pre-geminate lengthening can also block otherwise productive high vowel devoicing between two voiceless consonants (Han 1994; Takeyasu 2012; see also Fujimoto, this volume).

similarly that vowels before voiceless singletons are on average 36.9ms while those before voiceless geminates are 53.4ms. Furthermore, some studies even found that in C_1VC_2V contexts, C_1 is longer when C_2 is a geminate than when C_2 is a singleton (Han 1994; Port, Dalby, and O'Dell 1987) (cf. Takeyasu 2012 who found the opposite, shortening pattern; Hindi shows the same lengthening pattern: Ohala 2007).

On the other hand, vowels that follow geminate/singletons show the reverse pattern: those that follow geminate consonants are shorter than those that follow singleton consonants (Campbell 1999; Han 1994; Hirata 2007; Idemaru and Guion 2008; Ofuka 2003). Han (1994) found the shortening of post-geminate vowels (and sometimes also the following word-final moraic nasals) by 9ms. In an acoustic study reported in Idemaru and Guion (2008), the mean duration of the following vowel is 63ms after geminates and 76ms after singletons. As explicitly noted by Hirata (2007), however, this difference in duration of the following vowels is less substantial and less consistent than the difference in the preceding vowel.

Finally, one may expect that Voice Onset Time (VOT) – an interval between the release of the closure and the onset of voicing of the following vowel – would be longer for geminate stops than for singleton stops, because longer closure would result in higher pressure build-up behind the stop occlusion. However, this expectation does not seem to hold: in Han (1994), VOT is slightly shorter for geminates than for singletons; in other studies (Hirata and Whiton 2005; Homma 1981), the relationship is inconsistent. See Kokuritsu Kokugo Kenkyūjo (1990) for the data on the intraoral air pressure rise in Japanese consonants, which indeed shows that geminates do not involve higher intraoral air pressure rise.

2.2.2 Other non-durational, acoustic correlates

Several studies have investigated other non-durational, acoustic correlates of a singleton-geminate contrast in Japanese. Their findings are summarized in Table 2.

As observed in Table 2, Japanese geminates are associated with various non-durational cues. Given that, in addition to the primary acoustic correlate of constriction duration, there are a number of acoustic cues that are associated with Japanese geminates, they cannot be merely characterized as "long consonants".

A remaining question therefore is how to represent Japanese geminates phonologically. Many possibilities exist in answer to this question, such as (i) double consonants (often assumed in phonemic representation/transcription), (ii) moraic consonants (Hayes 1989), (iii) a special /Q/ phoneme – or *sokuon* – as assumed in the traditional literature (e.g. Hattori 1984), or (iv) a special syllable concatenater (Fujimura and Williams 2008). This issue should continue to be discussed in relation to the phonological behavior of Japanese geminates (see Kawagoe, this volume), as well as to the theory of phonetic implementation of phonological representations.

Table 2: A summary of other, non-durational, acoustic correlates of Japanese geminates. Reference keys: F = Fukui (1978), I&G = Idemaru & Guion (2008), O = Ofuka (2003), K = Kawahara (2006a). See the original papers for the details of the measurement procedures

	Patterns	References
Intensity	– The mean intensity difference between the surrounding vowels is larger across geminates.	I&G, O
F0	– F0 drop (a correlate of a lexical accent – see Kawahara, this volume) is larger across geminates.	I&G, O, K
	– F0 falls toward geminates in unaccented disyllabic words.	F
F1	– F1 is lower after geminates.	K
Spectral tilt	– H1-A1 is smaller for vowels after geminates (i.e. vowels are creakier).	I&G

2.2.3 The search for invariance

One general research program in phonetics is the search for invariance (Stevens and Blumstein 1981). The issue addressed in this program is whether, for each phonological distinction, there exists any acoustic correlate that is invariant across phonological contexts, individual speakers, and speech styles, etc., and if so, what those invariant acoustical properties are. This issue is particularly important for a singleton-geminate contrast, because, although geminates are longer than singletons given the same speech rate, geminates in fast speech styles can be shorter than singletons in slow speech styles (Hirata and Whiton 2005; Idemaru and Guion-Anderson 2010).[5]

Usually proposals for invariant measures take the form of a relationship between more than one acoustic parameter. The general idea behind these studies on phonological contrasts based on durations is rate normalization – listeners normalize the duration of incoming acoustic signal according to the speech rate, which can be (unconsciously) inferred from the duration of other intervals (Miller and Liberman 1979; Pickett and Decker 1960). For example, when a preceding vowel sounds short, a listener may perceive that the speaker is speaking fast, and as a result even a phonetically short interval may be interpreted as phonologically long.[6]

5 It has been observed in other languages (Italian: Pickett, Blumstein, and Burton 1999 and Persian: Hansen 2004) that geminates are more susceptible to change in duration due to speech rate than singletons are. This asymmetry seems to hold in the Japanese data as well (Hirata and Whiton 2005; Idemaru and Guion-Anderson 2010).

6 An alternative theory is auditory durational contrast. This auditory mechanism (more or less automatically) renders an interval to sound longer next to a shorter interval (this mechanism is sometimes referred to as "durational contrast"). This mechanism is arguably not specific to speech, as it applies to the perception of non-speech stimuli (Diehl and Walsh 1989; Kluender, Diehl, and Wright 1988). It is beyond the scope of this paper to compare these two theories (for further discussion on this debate, see Diehl, Walsh, and Kluender 1991; Fowler 1990, 1991, 1992; Kingston et al. 2009).

Several relational acoustic measures have been proposed as an invariant measure that distinguishes singletons from geminates across different speech rates. Hirata and Whiton (2005) recorded various disyllabic tokens of singletons and geminates in nonce words and real words in three speech styles (slow, normal, and fast), and considered three measures: (i) raw closure duration, (ii) C/V_1 ratio (the ratio between the target consonant and the preceding vowel), and (iii) C/W(ord) ratio. Hirata (2007) and Hirata and Forbes (2007) followed up on this study and considered three more measures: (i) C/V_2 ratio (V_2 = the following vowel), (ii) V-to-V interval (i.e. added durations of preceding vowel, constriction and VOT)[7] and (iii) VMora (V-to-V interval divided by average mora duration). Idemaru and Guion-Anderson (2010) tested yet a few more relational measures: C/V_1, C/C_1V_1, C/V_2, and $C/(C + V_2)$ (where C is the target consonant, C_1 and V_1 are the preceding consonant and vowel, and V_2 is the following vowel), in addition to those already tested by Hirata and Whiton (2005) (specifically, raw closure duration and C/W ratio). After recording their own various tokens of singletons and geminates in three speaking rates, for each measure, Idemaru and Guion-Anderson (2010) tested classification accuracy percentages based on raw values as well as *z*-transformed (normalized) values within each speaker. Finally, in the most recent study on this topic, Hirata and Amano (2012) introduced a yet new notion, subword, which is a disyllabic (C)V (C)CV sequence, which includes the target singleton and geminate consonant medially. This notion is equivalent to C/W in Hirata and Whiton's (2005) work, as they used only disyllabic words.

All of these studies used discriminant analyses for each proposed measure to calculate how many percentages of tokens are accurately classified as a member of the intended category. The classification accuracy percentages of all the measures in these studies are summarized in Table 3.

One tendency that is clear from Table 3 is that relational measures generally classify singletons from geminates better than raw durational values do. Just which relational measure best cross-classifies Japanese singletons from geminates is an interesting topic for on-going and future research. We cannot also deny the possibility that there are other measures, relational or not, which better cross-classify Japanese singletons and geminates, which are yet to be uncovered.[8]

Another important issue is the perceptual relevance – or reality – of the relational, invariant acoustic measures: whether Japanese listeners exploit relational, acoustic measures, and if so, which measures are they sensitive to. For example,

7 For example, given [kata], the V-to-V interval is [at], and given [katta], the V-to-V interval is [att].

8 Other relational invariant measures proposed for length contrasts in other languages include C/V_1 ratio for Italian (Pickett, Blumstein, and Burton 1999), vowel to rhyme duration ratio for Icelandic (Pind 1986) (in which long vowels and geminates are more or less in a complementary distribution), and the ratio of the closure duration to the syllable duration in Persian (with some further complications) (Hansen 2004).

Table 3: A summary of classification accuracy percentages in the five studies cited in the text (chronologically ordered). See text for explanations of each measure

Hirata and Whiton (2005)	
raw C duration	82.2% (nonce words) and 81.4% (real words)
C/V_1 ratio	92.1% (nonce words) and 91.3% (real words)
C/W	98% (nonce words) and 95.7% (real words)
Hirata (2007)	
C/V_2 ratio:	98.9% (nonce words) and 98.8-98.9% (real words)
Hirata and Forbes (2007)	
V-to-V interval	75.5%
VMora	99.6%
Idemaru and Guion-Anderson (2010)	
C/V_2	83.7% (raw) and 85.5% (normalized)
C/C_1V_1 (mora)	92.6% (raw) and 94.5% (normalized)
C/V_2	94.1% (raw) and 94.9% (normalized)
$C/(C+V_2)$	92.3% (raw) and 93.0% (normalized)
C/Word	96.3% (raw) and 96.8% (normalized)
raw C duration	87.2% (raw) and 88.3% (normalized)
Hirata and Amano (2012)	
C/W	97.5% (nonce words) and 93.9% (real words)
C/Subword (Subword=CV(C)CV)	97.6% (nonce words) and 93.6% (real words)

Idemaru and Guion-Anderson (2010) followed up their acoustic study with a perception test, which showed that while preceding mora (C_1V_1) duration significantly affects the perception of geminacy, whereas the following materials (C/V_2 ratio) do so only marginally, despite the fact that ratios involving these two factors yielded comparable accuracy percentages in production (see Table 3). See also Amano and Hirata (2010) and Otaki (2011) and section 3.2 for further discussion on the relationship between production and perception, especially in terms of contextual effects on the perception of length contrasts.

2.3 Manner and voicing effects

One issue that has received relatively less attention in the previous literature on Japanese geminates is the comparison of different manners of geminates in Japanese. Most previous acoustic studies on Japanese have investigated oral stops (Beckman 1982; Han 1992, 1994; Hirata and Whiton 2005; Hirose and Ashby 2007; Homma 1981; Idemaru and Guion 2008; Kawahara 2006a), although some studies included geminates of various manner types (e.g. Han 1962 measured oral stops and nasals; Campbell 1999 measured stops and some fricatives). Other languages that have been

Table 4: The effects of manner of articulation on the duration of singletons and geminates in Japanese (margin of error for 95% confidence intervals.)

Segment	Singleton	Geminate	Ratio
[p]	77.3 (7.8)	129.6 (8.1)	1.68
[t]	55.5 (4.6)	124.4 (7.3)	2.24
[k]	67.3 (7.1)	128.7 (7.1)	1.91
[b]	53.1 (3.8)	131.4 (8.8)	2.47
[d]	36.6 (1.9)	116.0 (10.4)	3.16
[g]	52.1 (3.7)	115.0 (13.2)	2.20
[ɸ]	83.5 (4.8)	144.7 (7.4)	1.73
[s]	83.2 (4.6)	134.5 (7.0)	1.62
[ʃ]	85.9 (5.7)	138.4 (7.3)	1.61
[ç]	63.4 (2.5)	132.0 (6.2)	2.08
[h]	72.2 (4.2)	143.7 (6.4)	1.99

studied in this light – manner effects on geminate contrasts – include Italian (affricates: Faluschi and Di Benedetto 2001; fricatives: Giovanardi and Di Benedetto 1998; nasals: Mattei and Di Benedetto 2000; see also Payne 2005), Cypriot Greek (Tserdanelis and Arvaniti 2001), Guinaang Bontok (Aoyama and Reid 2006), Finnish (Lehtonen 1970), Buginese, Madurese, and Toba Batak (Cohn, Ham, and Podesva 1999).

2.3.1 Fricative geminates

Japanese allows both (voiceless) stops and fricatives to contrast in geminacy. As in other languages (Lehiste 1970), singleton fricatives are generally longer than singleton stops in Japanese (Beckman 1982; Campbell 1999; Port, Dalby, and O'Dell 1987; Sagisaka and Tohkura 1984). As a result, geminate/singleton duration ratios are smaller for fricatives than for stops. Table 4 reports unpublished data collected by the author based on three female Japanese native speakers. All speakers were in their twenties at the time of recording, and the recording took place in a sound-attenuated room. Each target sound was pronounced in a nonce word frame [ni_o] (for most cases), itself being embedded in a frame sentence. Accents were always placed on the initial syllables. All three speakers repeated 10 repetitions of all tokens.[9]

Table 4 shows the results of duration measurements (for stops, VOT's were not included in the closure duration, as in many studies cited in Table 1). Duration ratios

9 I am grateful to Kelly Garvey and Melanie Pangilinan for their help with this acoustic analysis. This project also measured the duration of singleton and geminate nasals. The result shows that the geminate/singleton duration ratio for [n] was about 2.2 (Kawahara 2013a).

are highest for voiced stops than for voiceless stops (see also Homma 1981 and Hirose and Ashby 2007 for the same finding), which are also generally higher than for fricatives (except for [ç] and [h]).[10]

One phonological importance of the difference between stop pairs and fricative pairs is that the length contrast may be less perceptible for fricatives than for stops.[11] This less perceptible contrast of fricative pairs may lead to a diachronic neutralization (Blevins 2004) and/or avoidance of fricative geminates in synchronic phonological patterns (Kawahara 2006b, 2013b) due to a principle of contrastive dispersion to avoid contrasts that are not very well perceptible (Engstrand and Krull 1994; Flemming 2004; Liljencrants and Lindblom 1972; Lindblom 1986 and references cited therein; see also Martin and Peperkamp 2011 for a recent review on the effect of speech perception on phonological patterns.).

2.3.2 Affricate geminates

Affricates ([ts]) are not contrastive in the native phonology of Japanese, appearing as an allophonic variant of /t/ before [ɯ] (see Pintér, this volume); Geminate [ts], however, appears marginally in some borrowing as in [kjattsɯ] "cats" (see Kubozono, this volume). For this reason, the phonetic properties of affricate geminates have been much understudied. As far as I know, the only extensive study is that is offered by Oba, Brown, and Handke (2009), who found that the primary acoustic correlate of affricate geminates seems to lie in the difference in the closure duration, and not in frication duration. More studies on the properties of affricate geminates in Japanese are hoped for.

2.3.3 Voiced obstruent geminates

Finally, the effect of voicing on geminates is no less interesting. The native phonology of Japanese does not allow voiced obstruent geminates (Ito and Mester 1995, 1999; Kuroda 1965). The lack of voiced obstruent geminates has been argued to be due to their aerodynamic difficulty (Hayes and Steriade 2004; Ohala 1983; Westbury and

---

10 This study also found that the duration ratio of [p]-[pp] is smaller than that of [t]-[tt] and [k]-[kk]. This lower ratio may be related to the fact that length is not contrastive for [p] in the native phonology in Japanese (see Ito and Mester 1995, 1999 and Nasu, this volume). One puzzle, however, is why voiced stops have high duration ratios despite the fact that they are not contrastive in native Japanese phonology (Ito and Mester 1995, 1999). See also Engstrand and Krull (1994) for the relationship between the functional load of length contrasts and their phonetic realization. A full consideration on this relationship should be explored in future studies.

11 Whether there indeed is a difference in perceptibility between stops and fricatives should be tested in a perception study.

Keating 1986, and more references cited in Kawahara 2006a). For voiced stops, the intraoral air pressure goes up behind the oral stop closure; this rise in the intraoral air pressure makes it difficult to maintain the airflow required for vocal fold vibration. For voiced fricatives, the intraoral airpressure must rise to create frication, which again makes it difficult to maintain the transglottal air pressure drop. Perhaps for these reasons (synchronically or diachronically), the native phonology of Japanese does not allow voiced obstruent geminates.

However, gemination found in the context of loanword adaptation resulted in voiced obstruent geminates (e.g. Katayama 1998; Kubozono, Ito, and Mester 2008; Shirai 2002; see also Kawagoe, this volume, and Kubozono, this volume); e.g. [heddo] 'head' and [eggɯ] 'egg'. Nevertheless, presumably due to the aerodynamic difficulty, voiced geminate stops are generally "semi-devoiced" in Japanese. All three speakers recorded in Kawahara (2006a) show semi-devoicing. Figures 3 and 4 illustrate the difference between singletons and geminates: for singleton [g], closure voicing is fully maintained, while for geminate [gg], voicing during the stop closure ceases in the middle of the whole closure.[12] In Kawahara (2006a), on average, voicing is maintained only about 40% of the whole closure. Hirose and Ashby (2007) replicate this finding, showing that voiced Japanese geminates have only 47% of closure voicing.

As far as I know, there is no quantitative study on the phonetic implementation on voiced geminate fricatives in Japanese – this is a topic which is worth pursuing in a future study.[13]

One notable aspect of this semi-devoicing of geminates is that the following word-final high vowels after "semi-devoiced" geminates (e.g. [egg̊ɯ] 'egg') do not devoice, even though the vowels are preceded by a – phonetically speaking – voiceless interval (Hirose and Ashby 2007). The lack of high vowel devoicing in this context shows that the (semi-devoiced) voiced geminates are still phonologically voiced, and that high vowel devoicing is conditioned by phonological, rather than, phonetic factors. See Fujimoto (this volume) for further discussion on this debate.

The semi-devoicing of voiced obstruent geminates is found in other languages (e.g. (Tashlhiyt) Berber: Ridouane 2010), but it is not universal, despite the fact that it presumably arises from a physical, aerodynamic difficulty (Ohala 1983). Cohn, Ham and Podesva (1999) show, for example, that Buginese, Madurese, and Toba Batak all maintain voicing throughout the geminate closure; Egyptian Arabic is another language which has fully voiced geminates (Kawahara 2006a), and Lebanese

12 These spectrograms are based on new recordings made for Kawahara (2013c).

13 Voiced fricatives in Japanese become affricates word-initially, although whether this alternation is in free-variation or an allophonic alternation is controversial (Maekawa 2010). Osamu Fujimura (p.c., April 2012) points out that this hardening process may also happen when voiced fricatives become geminates as well, as in [oddzɯ] 'odds'. Affrication process may then be a general hardening process, which occurs in phonetically strong positions (i.e. word-initially and in geminates).

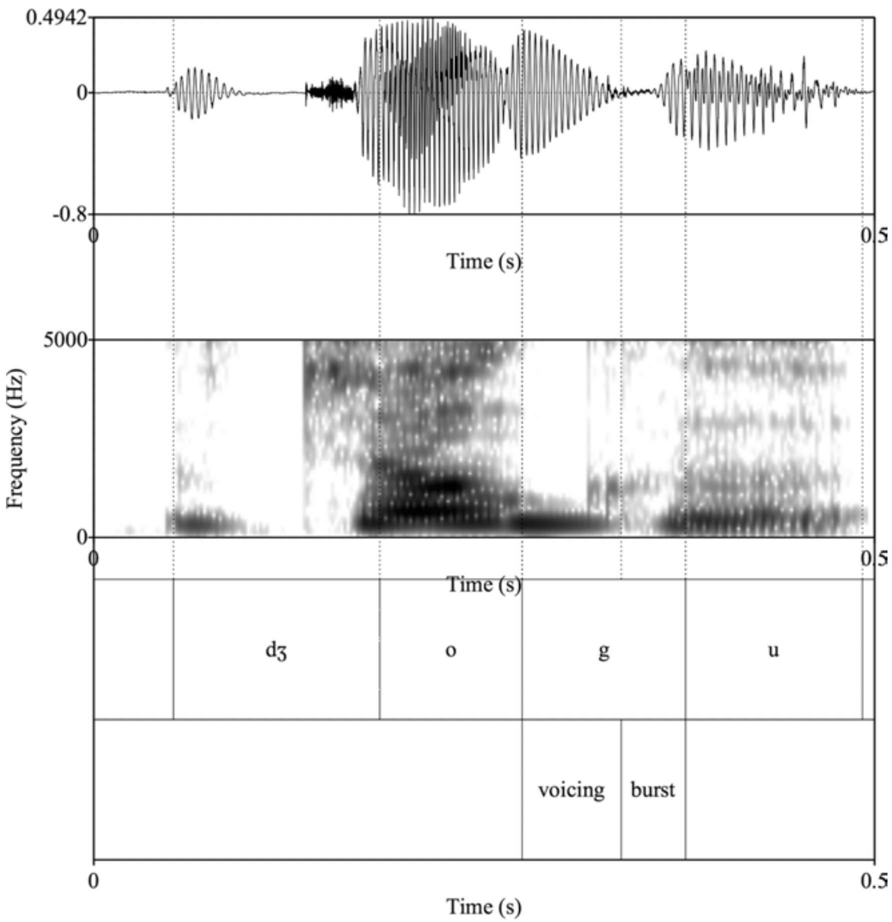


Figure 3: A singleton [g]

Arabic shows high percentages of voicing maintenance in medial, non-final positions (Ham 2001). Several Japanese dialects in Kyushu, including the Nagasaki dialect, also seem to show fully voiced geminate stops (Matsuura 2012). Cohn, Ham, and Podesva (1999) speculate that speakers resort to extra articulatory maneuvers like larynx lowering and cheek expansion to deal with the aerodynamic challenges (Ohala 1983). These articulatory gestures expand the size of oral cavity, thereby lowering the intraoral pressure (by Boyle's Law), providing the sufficient transglottal air pressure drop necessary to maintain vocal fold vibration (see Hayes and Steriade 2004, Ohala 1983, Ohala and Riordan 1979, and others).

The reason that (non-Kyushu) Japanese speakers do not deploy such articulatory strategies – at least not to the extent that geminates are fully voiced – may be that

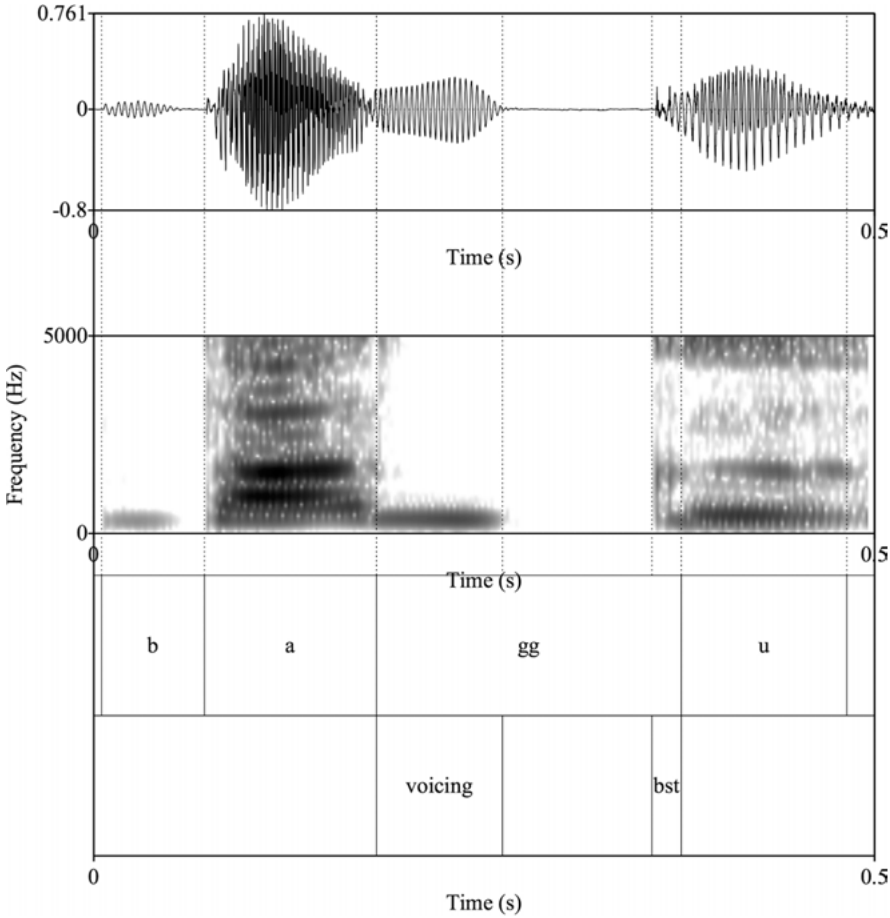


Figure 4: A geminate [gg]

voiced obstruent geminates are historically relatively new (see Pintér, this volume), and therefore the functional load of a voicing contrast in geminates is low, the contrast being restricted to loanwords (Ito and Mester 1995, 1999); i.e. there are not many minimal pairs. It would thus be interesting to observe whether speakers of future generations would start producing fully-voiced geminates, if the voicing contrast in geminate becomes more widespread in the Japanese lexicon. Moreover, a further cross-linguistic comparison is warranted to explore the relationship between how voiced stop geminates are implemented, and how the particular phonetic implementation patterns affect their phonological patterns (if they do at all) (see Kawahara 2006a for discussion).

2.4 Comparison with other languages

2.4.1 Constriction duration

I have already mentioned a few differences and similarities between Japanese geminates and geminates found in other languages, but now we turn our attention to a more detailed comparison of Japanese with other languages. As reviewed in section 2.1, Japanese geminates are acoustically characterized by long constriction duration, almost always twice as long as corresponding singletons. Similarly, constriction duration is usually the primary acoustic correlate of a singleton/geminate contrast in other languages; e.g. (Lebanese) Arabic (Ham 2001), Bengali (Lahiri and Hankamer 1988), Berber (Ridouane 2010), Bernese (Ham 2001), Buginese (Cohn, Ham, and Podesva 1999), Estonian (Engstrand and Krull 1994), Finnish (Engstrand and Krull 1994; Lehtonen 1970), Cypriot Greek (Tserdanelis and Arvaniti 2001), Guinaang Bontok (Aoyama and Reid 2006), Hindi (Ohala 2007; Shrotriya et al. 1995), Hungarian (Ham 2001), Italian (Esposito and Di Benedetto 1999; Payne 2005; Pickett, Blumstein, and Burton 1999), Madurese (Cohn, Ham, and Podesva 1999), Malayalam (Local and Simpson 1999), Pattani Malay (Abramson 1987b), Persian (Hansen 2004), Swedish (Engstrand and Krull 1994), Swiss German (Kraehenmann and Lahiri 2008), Toba Batak (Cohn, Ham, and Podesva 1999), and Turkish (Lahiri and Hankamer 1988) (see Kawahara and Braver 2014 and Ridouane 2010 for more languages and references).

One interesting cross-linguistic difference is the size of duration ratios between singletons and geminates. In Norwegian, for example, the ratio is much smaller than in Japanese (ranging from 1.22–1.38 in medial positions, cf. Table 1), and more substantial differences manifest themselves in the duration of preceding vowels (Fintoft 1961) (although one should note that Fintoft measured only non-stop consonants; see section 2.3.1).[14] In Buginese and Madurese, the geminate/singleton duration ratios are generally below 2 (Cohn, Ham, and Podesva 1999). Generalizing this observation, Ham (2001) entertains the possibility that geminate/singleton duration ratios are smaller for syllable-timed languages (e.g. Norwegian) than for mora-timed languages (e.g. Japanese). See also Maekawa (1984) for a comparison between Standard Tokyo dialect and Akita dialect – a dialect that has been described as syllable-timed – which points to the same generalization.

2.4.2 Other durational correlates

As discussed in section 2.2.1, vowels are longer before geminates in Japanese. This observation may come as a surprise given a cross-linguistic tendency that vowels in

14 Accordingly, when perceiving a singleton/geminate contrast, Norwegian speakers substantially rely on preceding vowel duration, much more than speakers of other languages (Kingston et al. 2009).

closed syllables are often shorter than vowels in open syllables (Maddieson 1985). Indeed many languages have shorter vowels before geminates than before singletons; e.g. Bengali (Lahiri and Hankamer 1988), Berber (Ridouane 2010), Italian (Esposito and Di Benedetto 1999; Pickett, Blumstein, and Burton 1999), Hindi (Ohala 2007; Shrotriya et al. 1995), Malayalam (Local and Simpson 1999), and the three Polynesian languages studied by Cohn, Ham, and Podesva (1999).

However, there are other languages that arguably show lengthening of vowels before geminates: Turkish,[15] Finnish (Lehtonen 1970, pp. 110–111), Shinhala (Letterman 1994) (although only one of the two speakers showed clear evidence) and Persian (Hansen 2004) (although no direct statistical tests are reported). The existence of such languages shows that Japanese may not simply be a typological anomaly, but languages vary in whether geminates shorten or lengthen the preceding vowels. I will come back to this issue of this cross-linguistic difference in section 3.2 in relation to its perceptual relevance.

In some languages, there are no substantial differences in the preceding vowel duration with singletons and geminates; e.g. Egyptian Arabic (Norlin 1987), Lebanese Arabic (at least for short vowels) (Ham 2001), Estonian (Engstrand and Krull 1994), and Hungarian (Ham 2001). In Cypriot Greek, there is slight tendency toward shortening before geminates, but this tendency is not very consistent (Tserdanelis and Arvaniti 2001).

Finally, the lack of an effect of geminacy on VOT in Japanese is paralleled in many languages including Buginese, Madurese, Toba Batak (Cohn, Ham, and Podesva 1999), Bernese, Hungarian, Lebanese Arabic (Ham 2001), Bengali (Hankamer, Lahiri, and Koreman 1989), and Berber (Ridouane 2010). Cypriot Greek has consistently longer VOT for geminates (Tserdanelis and Arvaniti 2001), but Turkish shows shorter VOT for geminates (Lahiri and Hankamer 1988).

2.4.3 Other non-durational, acoustic correlates

In addition to the durational correlates, different languages seem to show different non-durational acoustic correlates to signal singleton-geminate contrasts. These non-durational correlates are summarized in (1)–(6).[16]

(1) Bengali (Hankamer, Lahiri and Koreman 1989)
   a. Root Mean Square (RMS) amplitude of the following syllable is higher after singletons.

15 In Lahiri and Hankamer (1988), the difference is small and not statistically significant; see also Jannedy (1995) for evidence that this lengthening applies to closed syllables in general, as in Japanese (see footnote 4).

16 See the original references for stimulus designs and measurement procedures.

(2) Berber (Ridouane 2010)
   a. Geminates have higher amplitude during release.
   b. Geminates show burst release more consistently than singletons.

(3) Hindi (Shrotriya et al. 1995)
   a. F0 rises toward geminates in the preceding vowel.
   b. Burst intensity is stronger for geminates (by about 10dB).

(4) Italian (Payne 2006, based on electropalatographic (EPG) data)
   a. Geminates involve a more palatalized constriction than singletons.
   b. Geminate stops involve a more complete occlusion.
   c. Geminates are associated with a laminal gesture; singletons are associated with an apical gesture.

(5) Malayalam (Local and Simpson 1999)
   a. Sonorant geminates show palatal resonance with higher F2.
   b. The surrounding vowels differ in F1 and F2.

(6) Pattani Malay
   a. The peak amplitude of initial vowels (with respect to the following vowel) is higher after word-initial geminates than singletons (Abramson 1987b, 1998).
   b. Fundamental frequency of word-initial vowels is higher after word-initial geminates (Abramson 1998).
   c. First vowels are longer (with respect to second vowels) after word-initial geminates (Abramson 1998).
   d. The slope of amplitude rise is steeper after word-initial geminates (Abramson 1998).

So far Idemaru and Guion (2008) is the most extensive study looking for spectral correlates of geminacy contrasts in Japanese, and it is yet to be investigated whether the correlates listed in (1)–(6) are found in Japanese (though the Pattani Malay case may be special because it involves cases of word-initial geminates). However, it seems likely at this point that the phonetic implementation patterns of singleton-geminate contrasts are language-specific, the only universal rule being that geminates are longer than singletons (Ham 2001; Ridouane 2010). A remaining task in the phonetic theory is how to model the universality and language-specificity of phonetic implementation patterns of length contrasts. We should also perhaps bear in mind that "geminates" in different languages may not be the same phonological entity – there remains a possibility that these "geminates" have different phonological representations. See also Davis (2011) for relevant discussion.

3 The perception of geminates in Japanese

We now turn to the perception of a singleton-geminate contrast, beginning with a discussion of cues used by Japanese listeners and continuing with a discussion of cross-linguistic cues for geminacy contrasts.

3.1 The primary cue: constriction duration

Many studies have shown that the longer the constriction, the more likely the target is perceived as a geminate. This effect has been shown to hold in many perception studies using Japanese listeners (Amano and Hirata 2010; Arai and Kawagoe 1998; Fujisaki, Nakamura, and Imoto 1975; Fujisaki and Sugito 1977; Fukui 1978; Hirata 1990; Kingston et al. 2009; Oba, Brown, and Handke 2009; Takeyasu 2012; Watanabe and Hirato 1985). As an example, Figure 5 reproduces the results of Kingston et al.

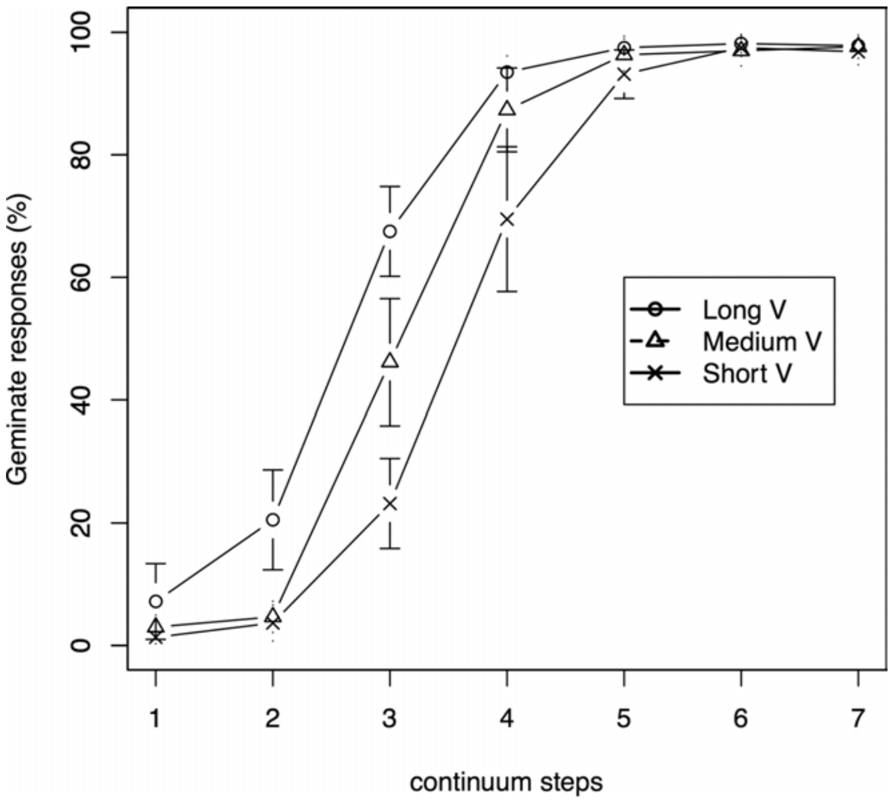


Figure 5: The effect of closure duration and the preceding vowel duration on the perception of geminates by Japanese listeners. Adapted from Kingston et al. (2009). Reprinted with permission from Elsevier

(2009) in which closure duration was varied from 60ms and 150ms in 15ms increments (see the next section for the three vocalic contexts). We observe that geminate responses increase as closure duration increases.

3.2 Contextual effects

More controversial than the effects of constriction duration are contextual effects. Fukui (1978) found that when the closure duration of an original singleton consonant was lengthened, it was almost always perceived as a geminate when the closure duration was doubled. On the other hand, shortening an original geminate did not result in a comparable shift in perception. The results show that closure duration is not the only cue for perceiving geminates. Similar types of effects (albeit to different degrees) were found in similar types of experiments on other languages (Bengali: Hankamer, Lahiri, and Koreman 1989, Pattani Malay: Abramson 1987a, 1992, Tamil: Lisker 1958, and Turkish: Hankamer, Lahiri, and Koreman 1989).

As reviewed in section 2.2.1, given that vowels are longer before geminates, we expect that Japanese speakers are more likely to perceive a consonant as a geminate after a longer vowel than after a shorter vowel. Several results indeed found a contextual effect in this direction (Arai and Kawagoe 1998; Kingston et al. 2009; Ofuka 2003; Ofuka, Mori, and Kiritani 2005; Takeyasu 2012). This contextual effect is illustrated in Figure 5 in which listeners judged more of the continuum as geminates after longer vowels.

On the other hand, several studies have found opposite results as well. For example, Watanabe and Hirato (1985) found that the perceptual boundaries between singletons and geminates shift toward longer duration after longer vowels, meaning that longer duration was required after longer vowels for consonants to be perceived as geminates (although only two listeners participated in this study). A similar boundary shift was found in Hirata (1990). Idemaru and Guion-Anderson (2010) kept the duration of the consonant at about 140ms and changed the duration of the preceding mora ($C_1 + V_1$ = onset plus preceding vowel), and found that the shorter the preceding mora duration, the more geminate responses were obtained. On the other hand, Takeyasu (2012) argues that it is the duration of C_1/V_1 ratio that matters, and that higher C_1/V_1 ratios lead to more geminate percepts. For more references of studies that obtained the results in this direction, see also Fujisaki and Sugito (1977)[17] and Idemaru, Holt, and Seltman (2012).

In summary, some studies found an "assimilative" pattern (more geminate responses after longer vowels) while others found a "contrastive" pattern (more geminate responses after shorter vowels). Where the difference between the two types

17 Fujisaki and Sugito (1977) found a contextual effect for the /s/-/ss/ contrast, but the paper is not explicit about the other two geminate pairs (/t/-/tt/ and /m/-/mm/).

of results comes from is an interesting question. There is some evidence that the magnitudes of the duration ratios between the target and context may matter in this regard (Nakajima, ten Hoopen, and Hilkhuysen 1992). Takeyasu (2012) also entertains the hypothesis that in experiments that obtained an contrastive effect, listeners may have judged the preceding vowels to be phonologically long, in which case the listeners are biased against judging the following consonant as long to avoid a superheavy syllable (see Ito and Mester, this volume, and Kubozono 1999 for a phonological constraint against superheavy syllables in Japanese, and Kawagoe and Takemura 2013 for its perceptual impact). Further experimentation is necessary to settle this issue.

Unlike preceding vowels, vowels are shorter after geminates than after singletons (Campbell 1999; Han 1994; Idemaru and Guion 2008; Ofuka 2003) (see section 2.2.1). While Hirato and Watanabe (1987) found no effects of the duration of the following vowel on the perception of geminates, Ofuka, Mori and Kiritani (2005) did in fact find that listeners are more likely to judge stimuli as a geminate before a shorter vowel; Idemaru and Guion-Anderson (2010) found a similar contextual effect of following vowels, although they found the effect of preceding C_1V_1 mora to be more substantial. See also Nakajima, ten Hoopen, and Hilkhuysen (1992) for a relevant discussion.

Another issue is the (non-)locality of contextual effects. For example, Hirata (1990) tested the effect of sentence level speech rate on perception of length contrasts, and found that the duration of the whole sentential materials following the target word can impact the perception of geminates. The study found that those tokens which are unambiguously identified as either a singleton or a geminate can be perceived as a member of a different category if the following materials provide enough cues for speech rate.

When listeners normalize the perceived duration for speech rate, one remaining question is: to what extent do they rely on local cues like the immediately preceding/following vowels or (CV) moras or (C)VC(C)V subword (Hirata and Amano 2012), and to what extent do they rely on more global cues (like the entire word or utterance). On the one hand, in terms of psycholinguistic computational simplicity, local cues are presumably easier to track (Idemaru and Guion-Anderson 2010). Nevertheless, some studies (Amano and Hirata 2010; Hirata 1990; Pickett and Decker 1960) show the effect of global cues; for example, by comparing several relational measures, Amano and Hirata (2010) demonstrate that the relationship between consonant duration and entire word duration[18] provides a good perceptual cue to a length distinction in Japanese. Recall also that Hirata (1990) found contextual effects at sentential levels.

---

18 They demonstrate that it is not a simple ratio between these two durations, but a regression function with an intercept that most accurately predicts the perceptual behavior of Japanese listeners. This function is equivalent to the ratio between closure duration (c) plus some constant (k) and word duration (w); i.e. $(c + k)/w$.

However, taking into account a whole word or sentence to determine a length property of a singleton/geminate contrast may impose a psycholinguistic burden. In order to identify what the word is, it is necessary to determine whether the consonant in question is a singleton or a geminate, but in order for listeners to determine whether the consonant is singleton or a geminate, they need to know what the word is – there may be a chicken-and-egg problem here.

I do not wish to imply that this challenge is insurmountable, rather that more phonetic and psycholinguistic research is necessary to address this issue. Hirata (2007) suggests that gating experiments (Grosjean 1980) may address the issue of the (non-)locality of the perception of length contrasts. In this way, the relationship between production and perception of geminates in Japanese (as well as in other languages) provides an interesting forum of research, which may bear on the general theory of speech perception (see Amano and Hirata 2010, Hirata and Amano 2012, Idemaru and Guion-Anderson 2010, Idemaru, Holt, and Seltman 2012, Otaki 2011, Pind 1986: and others for discussion).

Another remaining question is how non-durational cues – F0 values and movement, spectral envelope, burst intensity, etc. (see also Table 2) – interact with durational cues in the perception of Japanese geminates. For example, Ofuka (2003) observes that geminates are shorter in accented disyllabic words than in corresponding unaccented words, and also that in perception, a consonant with a particular duration is more likely to be perceived as a geminate when the word is accented (see also Hirata 1990 who obtained similar results). Likewise, Kubozono, Takeyasu, and Giriko (2013) show that English monosyllabic utterances with falling pitch contours – which are acoustically similar to Japanese pitch accents (Kawahara, this volume) – are more likely to be perceived as geminates by Japanese listeners. On the other hand, Idemaru (2011) did not find any substantial effects of amplitude or the steepness of F0 fall on the perception of geminacy for Japanese listeners. More extensive studies are warranted to investigate the intricacy of perception of geminates in Japanese.

3.3 Comparison with other languages

Like Japanese, the effect of constriction duration on the perception of duration has been found in many languages; e.g. Arabic (Obrecht 1965), Bengali (Hankamer, Lahiri, and Koreman 1989), English[19] (Pickett and Decker 1960), Finnish (Lehtonen 1970), Hindi (Shrotriya et al. 1995), Italian (Esposito and Di Benedetto 1999; Kingston et al. 2009), Norwegian (Kingston et al. 2009), Pattani Malay (Abramson 1987a, 1992), and Turkish (Hankamer, Lahiri, and Koreman 1989).

19 English does not have a lexical geminate contrast; this experiment tested a pair like *topic* vs. *top pick* where one member of the pair contains multiple morphemes.

Across languages, the effect of a language particular phonetic implementation pattern – shortening or lengthening of the preceding vowel – is often reflected in the perception pattern as well. For example, unlike in Japanese, in both Norwegian (Fintoft 1961) and Italian (Esposito and Di Benedetto 1999), vowels are shorter before geminates. This shortening affects the perception of geminates – listeners of these languages are more likely to perceive a consonant as a geminate before a shorter vowel than a longer vowel (Esposito and Di Benedetto 1999; Kingston et al. 2009; van Dommelen 1999). In Icelandic, in which long vowels and geminates are in a complementary distribution, Pind (1986) shows that vowel duration with respect to the entire rhyme duration is a good predictor of geminate perception – given fixed rhyme durations, shorter vowel durations yielded more geminate responses.

One interesting puzzle that arises from this cross-linguistic comparison regarding shortening vs. lengthening in pre-geminate position is as follows: some researchers propose that C/V duration ratios provide mutually enhancing perceptual cues for duration when a shorter consonant is preceded by a longer vowel, as is the case for voicing contrasts in many languages (Kingston and Diehl 1994; Kohler 1979; Pickett, Blumstein, and Burton 1999; Port and Dalby 1982). A combination of a short vowel and a long consonant yields enhanced, high C/V_1 duration ratios, whereas a combination of a long vowel and a short consonant yields low ratios. Languages like Italian and Norwegian, in which preceding vowels are shorter before geminates, can be assumed to deploy this perceptual enhancement pattern. In this light, a question arises why Japanese lengthens a vowel before a geminate.

A tentative answer that I can offer is that V_1C unit (or V-to-V interval) may constitute another kind of perceptual unit, a unit that has been hypothesized to play a role in the perception of Japanese and other languages (Hirata and Forbes 2007; Kato, Tsuzaki, and Sagisaka 2003; Kingston et al. 2009; Ofuka, Mori, and Kiritani 2005; Sato 1978; van Dommelen 1999).[20] If V_1C is an important perceptual unit – whether it is universal or specific to Japanese – then a longer vowel before a geminate can be considered as perceptually enhancing the longer duration of geminates.

4 The articulatory characteristics of Japanese geminates

Compared to acoustic and perception studies of Japanese geminates, there are relatively fewer studies on the articulation of Japanese geminates, although there

20 An alternative idea is that although Japanese is a mora-timed language (where a mora usually constitutes a CV unit), geminates, whose coda part should constitute its own mora, are not by themselves as long as a CV unit; pre-geminate vowel lengthening may occur to compensate for this shortage of duration, as hypothesized and discussed by Warner and Arai (1999). See also Otake (this volume) for more on mora-timing in Japanese. One puzzle for this explanation is why, then, Japanese speakers shorten the following vowels after geminates.

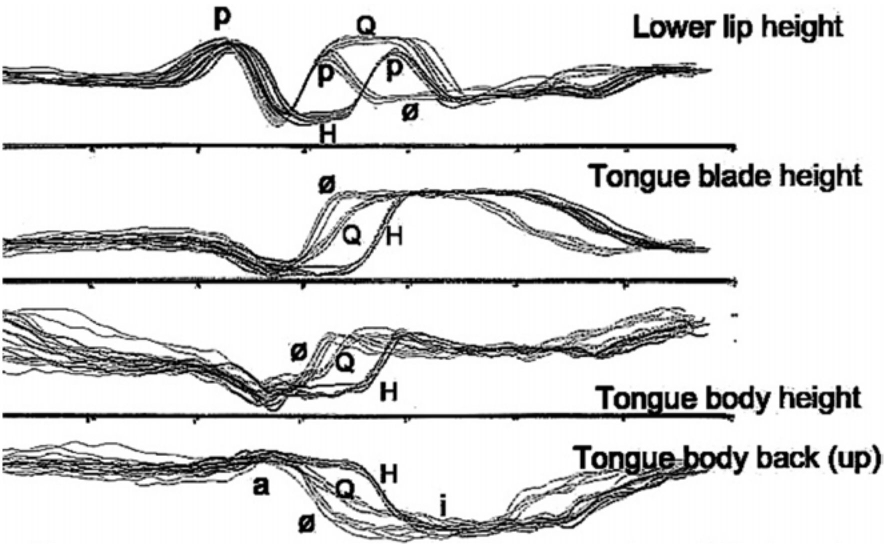


Figure 6: The articulatory movements of Japanese geminates, as compared to singletons. Based on Ishii (1999), cited and discussed in Fujimura & Williams (2008). Three conditions are [papa] (ϕ), [paapa] (H), and [pappa] (Q). Reprinted with permission from the author and the publisher

are some notable studies. Ishii (1999), for example, obtained articulatory data of Japanese geminates and long vowels using X-ray microbeam measurements, as shown in Figure 6. The three types of the stimuli were tested in this study, which were [papa] (ϕ), [paapa] (H), and [pappa] (Q).

Based on Figure 6, Fujimura and Williams (2008) make three observations. First, as we can observe in the top panel, a geminate [pp] in Japanese shows a prolonged lip closure compared to a singleton [p]. Second, while the lip movement toward its closure is comparable between singletons and geminates (the top panel) (though cf. Löfqvist 2007; Smith 1995), the lingual (tongue) movements are slower for geminates than for singletons (the second and the third panel). Finally, the V-to-V movement is slower and more gradual across geminates than across singletons (the bottom panel).

These results are corroborated by studies by Löfqvist (2006, 2007) using a magnetometer system. Longer constriction duration was confirmed for labial (Löfqvist 2006)[21] as well as alveolar and velar stops (Kochetov 2012; Löfqvist 2007). The speed of the tongue movement was found to be slower for alveolar and velar geminate stops than corresponding singletons (Löfqvist 2007). Slower V-to-V movement across geminate stops was also found by Löfqvist (2006).

21 Löfqvist (2006) studies nasal geminates, and therefore this finding is technically for *hatsuon*, not for *sokuon*.

Takada (1985) investigated X-ray data of Japanese consonants, and found two differences between singletons and geminates: slower movement in terms of lingual contact and jaw contact, with maximal contact formed at a later phase in the constriction in geminates. Smith (1995), again based on X-ray microbeam data, shows that a singleton/geminate distinction affects the gestural timing of the following vowel in Japanese, whereas in Italian it does not – she attributes this difference to differences in gestural coordination of vowels and consonants in Japanese and Italian. The EPG data by Kochetov (2012) shows greater degree of linguopalatal contact for geminates than for singletons. Sawashima (1968), using a fiberscope, shows that glottal abduction is larger for geminate fricatives than singleton fricatives. Finally, Kokuritsu Kokugo Kenkyūjo (1990) offers detailed articulatory data of Japanese sounds in general, including those of geminates.

5 Remaining issues

Although I have raised a number of remaining questions already, I would like to close this chapter with a discussion about several more questions that require further experimentation.

5.1 Non-intervocalic geminates

For lexical contrasts, Japanese allows geminates only intervocalically. However, some word-initial geminates are found due to an elision process in casual speech; e.g. [ttakɯ] from /mattaku/ (a phrase that often accompanies a sigh) and [sseena] from /usseena/ 'shut up'. Cues to word-initial geminates have been studied in some other languages (Abramson 1992, 1999; Kraehenmann and Lahiri 2008; Kraehenmann 2011; Muller 2001; Ridouane 2010), but the Japanese case has not been extensively investigated. A specific question is whether such word-initial geminates involve longer constriction just like intervocalic geminates. Articulatory studies, using devices like EPG (Kraehenmann and Lahiri 2008; Payne 2006; Ridouane 2010), would address the question of whether geminates do indeed involve a longer constriction word-initially (see Kraehenmann and Lahiri 2008 and Ridouane 2010 who found a positive answer to this question in Swiss German and Berber).

Similarly, an orthographic marker for Japanese geminates – "small tsu" – can also appear word-finally, especially in mimetic words (see Nasu, this volume), although this word-final gemination diacritic does not convey a lexical contrast. The exact nature of its phonetic realization is yet to be explored – impressionistically, it is realized as a glottal stop, but as far as I know, it has not been fully explored in instrumental work.

5.2 Derived geminates vs. underlying geminates

Some phonetic studies in other languages have compared lexical geminates and geminates derived by some phonological processes, most often by assimilation.[22] They have generally shown that lexical geminates and geminates derived by phonological processes are phonetically identical, as in Bengali (Lahiri and Hankamer 1988), Berber (Ridouane 2010), Sardinian (Ladd and Scobbie 2003), and Turkish (Lahiri and Hankamer 1988). However, Ridouane (2010) found a difference between lexical geminates and geminates created via morpheme concatenation in terms of preceding vowel duration and burst amplitude. Similarly, Payne (2005) argues that in Italian lexical geminates tend to be longer than post-lexical geminates created by RADDOPPIAMENTO SINTATTICO (RS) (although there are some complicating factors; see Payne 2006 for further discussion).

As far as I know, no studies have compared underlying and derived geminates in Japanese. For example, the final consonant of a prefix /maC-/ 'truly' assimilates to the root-initial consonant, resulting in a geminate (e.g. [mak-ka] 'truly red', [mas-sakasama] 'truly reversed', and [mam-marɯ] 'truly round'). It would be interesting to investigate whether there is a difference between such derived geminates and underlying geminates. One reason why we may expect a difference is as follows. Monomoraic roots in Japanese can be lengthened to have a long vowel, when pronounced in isolation without a case particle (Mori 2002); however, duration ratios between these lengthened vowels and short vowels are smaller than the ratios between underlying long vowels and short vowels found in the previous research – i.e., that this lengthening pattern is only incompletely neutralizing (Mori 2002 compares her results with the data from Beckman 1982 and Hoequist 1982; Braver and Kawahara 2014 confirmed that there are differences in duration between lengthened vowels and underlying long vowels within one experiment). It would be particularly interesting if we find such an incomplete neutralization pattern (Port and O'Dell 1985 *et seq.*)[23] in the context of gemination.

5.3 The phonetics of emphatic geminates

Japanese deploys gemination to convey emphatic meanings (e.g. [kattai] 'very hard' from [katai] 'hard') (Aizawa 1985; Kawahara 2001, 2006b, 2013b). In terms of orthography, this gemination can be written with multiple signs of gemination ("small tsu")

---

22 In some languages, geminates arise via simple morpheme concatenation without a further phonological change (known as "fake geminates"); e.g. /pat̪ + t̪e/ → [pat̪t̪e] 'spread out (INFINITIVE)' in Bengali (Lahiri and Hankamer 1988). In Japanese, fake geminates rarely if ever arise because root-final consonants always assimilate to the following consonant anyway; i.e. fake geminates would not be distinguishable from assimilated geminates.

23 For an overview of incomplete neutralizations, see Braver (2013), Kawahara (2011) and Yu (2011).

(Aizawa 1985). It would be interesting to investigate to what extent such repetition of geminate diacritics is reflected in actual production (and for that matter, can be tracked in perception). This issue is partly addressed by Kawahara and Braver (2014). A production study shows that at least some speakers can make a six-way duration differences, given five degrees of emphatic consonants (and non-emphatic consonants). Other speakers showed a steady correlation between emphasis levels and duration. The articulatory and perceptual properties of these emphatic geminates should be investigated more in future research.

Furthermore, this emphatic gemination pattern can create otherwise unacceptable types of geminates, such as voiced obstruent geminates in native words and approximant geminates (Aizawa 1985; Kawahara 2001; Kawahara and Braver 2014). Together with the general phonetic properties of emphatic geminates, the phonetic realization of approximant geminates in Japanese, in particular, is understudied and yet to be investigated.

5.4 The laryngeal "tension" of geminates

Despite the studies mentioned in section 4, the exact articulatory nature of Japanese geminacy contrasts is yet to be fully explored. One particular issue concerns whether Japanese geminates involve laryngeal constriction or not. Impressionistically, Japanese geminates are sometimes conceived of as having an accompanying glottal constriction. Hattori (1984) suggests that the first half of geminates involves glottal tension (p. 139). Aizawa (1985) uses a term "choked consonant" to refer to (emphatic) geminates. Idemaru and Guion (2008) also found shallower spectral tilt (H1-A1) in the vowels following geminates, indicating some creakiness, which implies some glottal constriction (although two other measures of creakiness did not show differences in their study). Fujimura and Williams (2008) argue that laryngealization is a distinctive characteristic of Japanese geminates, which may even contribute to the perception of geminates.

On the other hand, a study by Fujimoto, Maekawa, and Funatsu (2010) using a high-speed digital video recording system, did not find evidence for laryngeal or glottal tension in Japanese geminates. They also found that glottal opening is slightly larger during (voiceless) geminates than during singletons. Therefore, whether Japanese geminates involve glottal tension, and if so how that glottalization is coordinated/synchronized with supralaryngeal (oral) gestures, is still to be explored.

5.5 Dialectal differences

There are few cross-dialectal studies on Japanese geminates, especially those written in English, which would be available to those scholars who do not read the Japanese

literature. Due to the limitation of my expertise, I cannot discuss this issue extensively, but it would be particularly interesting to compare the properties of geminates in mora-timed dialects with syllable-timed dialects, such as the Aomori dialect (Takada 1985), the Akita dialect (Maekawa 1984), and the Kagoshima dialect (Kubozono and Matsui 2003).

5.6 Manner differences and the perception of geminates

Finally, as discussed in section 2.3, manner effects on the production of geminates in Japanese have been understudied. Relatedly, many perception experiments on Japanese geminates are based on voiceless stops (Amano and Hirata 2010; Arai and Kawagoe 1998; Hirata 1990; Hirato and Watanabe 1987; Fukui 1978; Idemaru and Guion-Anderson 2010; Kingston et al. 2009; Ofuka 2003; Takeyasu 2012; Watanabe and Hirato 1985). Fujisaki, Nakamura, and Imoto (1975) studied all manners, but nevertheless only report the results for fricatives (though see also Fujisaki and Sugito 1977 where they report the data for all manners). There are a few recent studies (Matsui 2012; Takeyasu 2009; Tews 2008), which investigated factors affecting the perception of geminates in fricatives. Oba, Brown, and Handke (2009) showed that the primary cue for affricate geminates lies in the closure phase, not in the frication phase. The production and the perception of different manners of geminates, including nasal geminates, warrants further investigation.

Acknowledgements

Thanks to Megan Moran, Mel Pangilinan, and Akiko Takemura for collecting many of the references cited herein. For discussion on this paper and/or comments on earlier versions of this draft, I thank Osamu Fujimura, Manami Hirayama, Michinao Matsui, Toshio Matsuura, Hope McManus, Kazu Kurisu, Akiko Takemura, and two reviewers (Itsue Kawagoe and Yukari Hirata). Hope McManus reformatted and proofread this paper. Some portions of this paper have appeared in Kawahara (2013a), which is written in Japanese and additionally discusses aspects of long vowels. A part of this research is supported by the Research Council Grant from Rutgers University. Remaining errors are mine.

References

Abramson, Arthur S. 1987a. The perception of word-initial consonant length: Pattani Malay. *Haskins Laboratories Status Report on Speech Research* SR-91. 149–156.

Abramson, Arthur S. 1987b. Word-initial consonant length in Pattani Malay. *Haskins Laboratories Status Report on Speech Research* SR-91. 143–147.

Abramson, Arthur S. 1992. Amplitude as a cue to word-initial consonant length: Pattani Malay. *Haskins Laboratories Status Report on Speech Research* SR-109/110. 251–254.

Abramson, Arthur S. 1998. The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length. In Udom Warotamasikkhadit and Thanyarat Panakul (eds.), *Papers from the Fourth Annual Meeting of the Southeast Asian Linguistic Society 1994*, Tempe, Arizona: Arizona State University, 1–20.

Abramson, Arthur S. 1999. Fundamental frequency as a cue to word-initial consonant length: Pattani Malay. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS 14), San Francisco: University of California, 591–594.

Abramson, Arthur S. and Leigh Lisker. 1985. Relative power of cues: F0 shift versus voice timing. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 25–33. Orlando: Academic Press.

Aizawa, Yoshiko. 1985. Intensification by so-called "choked sounds"–long consonants–in Japanese. *The Study of Sounds* (The Phonetic Society of Japan) 21. 313–324.

Amano, Shigeaki and Yukari Hirata. 2010. Perception and production boundaries between single and geminate stops in Japanese. *Journal of the Acoustical Society of America* 128(4). 2049–2058.

Aoyama, Katsura and Lawrence Reid. 2006. Cross-linguistic tendencies and durational contrasts in geminate consonants: An examination of Bontok geminates. *Journal of International Phonetic Association* 36. 145–157.

Arai, Masako and Itsue Kawagoe. 1998. Eigo no onsetsu-gata to sokuon chikaku – Nansensugo ni yoru chikaku tesuto no hōkoku [Consonant gemination and syllable types in English: A perceptual test of nonsense words]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 2(3). 87–92.

Beckman, Mary. 1982. Segmental duration and the 'mora' in Japanese. *Phonetica* 39. 113–135.

Behne, Darn, Takayuki Arai, Peter Czigler and Kirk Sullivan. 1999. Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS 14), San Francisco: University of California, 857–860.

Blevins, Juliette. 2004. Klamath sibilant degemination: Implications of a recent sound change. *International Journal of American Linguistics* 70. 279–289.

Braver, Aaron. 2013. *Degrees of incompleteness in neutralization: Paradigm uniformity with a phonetics with weighted constraints*. New Brunswick, New Jersery: Rutgers University dissertation.

Braver, Aaron and Shigeto Kawahara. 2014. Incomplete vowel lengthening in Japanese: A first study. In Santana-Labarge Robert E. (ed.), *Proceedings of WCCFL 31*. Somerville: Cascadilla Press, 86–95.

Campbell, Nick. 1999. A study of Japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan* 3(2). 29–39.

Cohn, Abigail, William Ham and Robert Podesva. 1999. The phonetic realization of singleton-geminate contrasts in three languages of Indonesia. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS 14), San Francisco: University of California, 587–590.

Davis, Stuart. 2011. Geminates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 873–897. Malden, MA and Oxford: Wiley-Blackwell.

Diehl, Randy and Margaret Walsh. 1989. An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America* 85. 2154–2164.

Diehl, Randy, Margaret Walsh and Keith Kluender. 1991. On the interpretability of speech/nonspeech comparisons: A reply to Fowler. *Journal of the Acoustical Society of America* 89(6). 2905–2909.

Engstrand, Olle and Diana Krull. 1994. Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica* 51. 80–91.

Esposito, Anna and Maria Gabriella Di Benedetto. 1999. Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America* 106. 2051–2062.

Faluschi, Simone and Maria Gabriella Di Benedetto. 2001. Acoustic analysis of singleton and geminate affricates in Italian. *WEB-SLS: The European Journal of Language and Speech* (online).

Fintoft, Knut. 1961. The duration of some Norwegian speech sounds. *Phonetica* 7. 19–39.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Rorbert Kirchner and Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Fowler, Carol. 1990. Sound-producing sources as the objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America* 88. 1236–1249.

Fowler, Carol. 1991. Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America* 89. 2910–2915.

Fowler, Carol. 1992. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics* 20. 143–165.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Fujimoto, Masako, Kikuo Maekawa and Seiya Funatsu. 2010. Laryngeal characteristics during the production of geminate consonants. In Takao Kobayashi, Keikichi Hirose and Satoshi Nakamura (eds.), *Proceedings of Interspeech 2010*, Chiba, 925–928.

Fujimura, Osamu and C. J. Williams. 2008. Prosody and syllables. *Phonological Studies* 11. 65–74.

Fujisaki, Hiroya, Kimie Nakamura and Toshiaki Imoto. 1975. Auditory perception of duration of speech and non-speech stimuli. In Gunnar Fant and Marcel A. A. Tatham (eds.), *Auditory analysis and perception of speech*, 197–219. London: Academic.

Fujisaki, Hiroya and Miyoko Sugito. 1977. Onsei no butsuriteki seishitsu [Physical natures of sounds]. In Susumu Ōno and Takesi Sibata (eds.), *Iwanami kōza nihongo 5: On'in* [Iwanami Japanese series 5: Phonology], 63–103. Tokyo: Iwanami.

Fukui, Seiji. 1978. Nihongo heisaon no enchō/tanshuku ni yoru sokuon/hisokuon to shite no chōshu [Perception for the Japanese stop consonants with reduced and extended durations]. *Onsei Gakkai Kaihō* [The Bulletin, The Phonetic Society of Japan] 159. 9–12.

Giovanardi, Maurizio and Maria Gabriella Di Benedetto. 1998. Acoustic analysis of singleton and geminate fricatives in Italian. *WEB-SLS: The European Journal of Language and Speech* (online).

Grosjean, Francois. 1980. Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics* 28(4). 267–283.

Ham, William. 2001. *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Han, Mieko. 1962. The feature of duration in Japanese. *Onsei no Kenkyū* [The Study of Sounds, The Phonetic Society of Japan] 10. 65–80.

Han, Mieko. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49. 102–127.

Han, Mieko. 1994. Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96. 73–82.

Hankamer, George, Aditi Lahiri and Jacques Koreman. 1989. Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal of Phonetics* 17. 283–298.

Hansen, Benjamin B. 2004. Production of Persian geminate stops: Effects of varying speaking rate. In Augustine Agwuele, Willis Warren and Sang-Hoon Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference*, 86–95. Somerville: Cascadilla Press.

Hattori, Shiro. 1984. *Onseigaku* [Phonetics]. Tokyo: Iwanami.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce and Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.), *Phonetically based phonology*, 1–33. Cambridge: Cambridge University Press.

Hirata, Yukari. 1990. Tango reberu bun reberu ni okeru nihonjin no sokuon no kikitori [Perception of geminated stops in Japanese word and sentence levels]. *Onsei Gakkai Kaihō* [The Bulletin, The Phonetic Society of Japan] 194. 23–28.

Hirata, Yukari. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32(4). 565–589.

Hirata, Yukari. 2007. Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels. *Journal of the Phonetic Society of Japan* 11(1). 9–22.

Hirata, Yukari. this volume. Chapter 18: L2 phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Hirata, Yukari and Connor Forbes. 2007. Vowel-onset interval as a timing unit for singleton/geminate stop distinction in Japanese. In Trouvain Jürgen and William J. Barry (eds.), *Proceedings of ICPhS XVI*, Saarbrücken, Germany, 349–354.

Hirata, Yukari and Stephen G. Lambacher. 2004. Role of word-external contexts in native speakers' identification of vowel length in Japanese. *Phonetica* 61. 177–200.

Hirata, Yukari and Kimiko Tsukada. 2009. Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* 66(3). 129–149.

Hirata, Yukari and Jacob Whiton. 2005. Effects of speaking rate on the singleton/geminate distinction in Japanese. *Journal of the Acoustical Society of America* 118. 1647–1660.

Hirata, Yukari and Shigeaki Amano. 2012. Production of singleton and geminate stops in Japanese three- and four-mora Japanese words. *Journal of the Acoustical Society of America* 132(3). 1614–1625.

Hirato, Nobuo and Shinichiro Watanabe. 1987. Sokuon no chikaku to kōzoku boin no jizoku jikan tono kankei [The relationship between the perception of moraic obstruents and the duration of the following vowels]. *Onsei Gengo* II. 99–106.

Hirose, Aki and Michael Ashby. 2007. An acoustic study of devoicing of the geminate obstruents in Japanese. In Trouvain Jürgen and William J. Barry (eds.), *Proceedings of ICPhS XVI*, Saarbrücken, Germany, 909–912.

Hoequist, Carles E. 1982. Durational correlates of linguistic rhythm categories. *Phonetica* 40. 19–31.

Homma, Yayoi. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9. 273–281.

Idemaru, Kaori. 2011. Role of amplitude and pitch in the perception of Japanese stop length contrasts. *Cross-cultural Studies* 24. 191–204.

Idemaru, Kaori and Susan Guion. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of International Phonetic Association* 38(2). 167–186.

Idemaru, Kaori and Susan Guion-Anderson. 2010. Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica* 67. 25–46.

Idemaru, Kaori, Lori Holt and Howard Seltman. 2012. Individual differences in cue weights are stable across time: The case of Japanese stops lengths. *Journal of the Acoustical Society of America* 132. 3950–3964.

Ishii, Toyota. 1999. Sokuon hatsuwaji no kōon kikan no undō ni kansuru kenkyū [A study of the movement of articulatory organs in Japanese geminate production]. *Journal of Otolarngology of Japan* 102. 622–634.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Oxford: Blackwell.

Ito, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Jannedy, Stephanie. 1995. Gestural phasing as an explanation for vowel devoicing in Turkish. *Ohio State University Working Papers in Linguistics* 45. 56–84.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology.* Santa Cruz, California: University of California dissertation.

Kato, Hiroaki, Minoru Tsuzaki and Yoshinori Sagisaka. 2003. Functional differences between vowel onsets and offsets in temproal perception of speech: Local-change detection and speaking-rate discrimination. *Journal of the Acoustical Society of America* 113(6). 3379–3389.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of sokuon, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawagoe, Itsue and Akiko Takemura. 2013. Geminate judgment of English-like words by Japanese native speakers: Differences in the borrowed forms of “stuff” and “tough”. *Journal of East Asian Linguistics* 22(4). 307–337.

Kawahara, Shigeto. 2001. Similarity among variants: Output-variant correspondence. Tokyo: International Christian University BA thesis.

Kawahara, Shigeto. 2006a. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82(3). 536–574.

Kawahara, Shigeto. 2006b. Mimetic gemination in Japanese: A challenge for Evolutionary Phonology. *Theoretical Linguistics* 32(1). 411–424.

Kawahara, Shigeto. 2011. Experimental approaches in theoretical phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 2283–2303. Malden, MA and Oxford: Wiley-Blackwell.

Kawahara, Shigeto. 2012. Amplitude changes facilitate categorization and discrimination of length contrasts. *IEICE Technical Report. The Institute of Electronics, Information, and Communication Engineers* 112. 67–72.

Kawahara, Shigeto. 2013a. Nihongo tokushu haku no onkyō to chikaku [The acoustics and perception of Japanese special moras]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 69(4). 1–6.

Kawahara, Shigeto. 2013b. Emphatic gemination in Japanese mimetic words: A wug-test with auditory stimuli. *Language Sciences* 40. 300–307.

Kawahara, Shigeto. 2013c. Testing Japanese loanword devoicing: Addressing task effects. *Linguistics* 51. 1271–1299.

Kawhara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto and Aaron Braver. 2013. The phonetics of emphatic vowel lengthening in Japanese. *Open Journal of Modern Linguistics* 3(2). 141–148.

Kawahara, Shigeto and Aaron Braver. 2014. The durational properties of emphatically-lengthened consonants in Japanese. *Journal of International Phonetic Association* 44(3). 237–260.

Kawahara, Shigeto, Melanie Pangilinan and Kelly Garvey. 2011. Spectral continuity and the perception of duration. Ms. Rutgers University. Talk presented at International Conference on Geminates 2011, Kobe University, 8–9 Jan.

Kingston, John and Randy Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kingston, John, Shigeto Kawahara, Della Chambless, Daniel Mash and Eve Brenner-Alsop. 2009. Contextual effects on the perception of duration. *Journal of Phonetics* 37(3). 297–320.

Kinoshita, Keisuke, Dawn Behne and Takayuki. Arai. 2002. Duration and F0 as perceptual cues to Japanese vowel quantity. *Proceedings of ICSLP*, 757–760.

Kluender, Keith, Randy Diehl and Beverly Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16. 153–169.

Kochetov, Alexei. 2012. Linguopalatal contact differences between Japanese geminate and singleton stops. *Canadian Acoustics* 40(3). 28–29.

Kohler, Klaus. 1979. Dimensions in the perception of fortis and lenis plosives. *Phonetica* 36. 332–343.

Kokuritsu Kokugo Kenkyūjo. 1990. *Nihongo no boin, shiin, onsetsu: Chōon undō no jikken onseigakuteki kenkyū* [Vowels, consonants and syllables in Japanese]. Tokyo: Shūei Shuppan.

Kraehenmann, Astrid and Aditi Lahiri. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *Journal of the Acoustical Society of America* 123(6). 4446–4455.

Kraehenmann, Astrid (2011) Initial geminates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1124–1146. Malden, MA and Oxford: Wiley-Blackwell.

Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo and Michinao Matsui. 2003. Phonetic vs. phonological control of speech: Closed syllable vowel shortening in Japanese dialects. In M. J. Solé, D. Recasens and J. Romero (eds.), *Proceedings of ICPhS XV*, 2429–2432.

Kubozono, Haruo, Junko Ito and Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. In The Linguistic Society of Korea (ed.), *Current issues in unity and diversity of languages. Collection of papers selected from the 18th International Congress of Linguists*, 953–973. Republic of Korea: Dongam Publishing.

Kubozono, Haruo (ed.). 2007. *Sokuon*, or moraic obstruent. *Journal of the Phonetic Society of Japan* 11(1).

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kubozono, Haruo, Hajime Takeyasu and Mikio Giriko. 2013. On the positional assymmetry of consonant gemination in Japanese loanwords. *Journal of East Asian Linguistics* 22(4). 339–371.

Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. Cambridge, MA: MIT dissertation.

Ladd, D. Robert and James Scobbie. 2003. External sandhi as gestural overlap? Counter-evidence from Sardinian. In John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in laboratory phonology VI: Phonetic interpretation*, 164–182. Cambridge: Cambridge University Press.

Lahiri, Aditi and George Hankamer. 1988. The timing of geminate consonants. *Journal of Phonetics* 16. 327–338.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: MIT Press.

Lehtonen, Jaakko. 1970. *Aspects of quantity in standard Finnish*. Jyvaeskylaessae: Osakeyhtioen kirjapainossa.

Letterman, Rebecca S. 1994. A phonetic study of Sinhala syllable rhymes. *Working Papers of the Cornell Phonetics Laboratory* 9. 155–181.

Liljencrants, Johan and Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John Ohala and Jeri Jaeger (eds.), *Experimental phonology*, 13–44. Orlando: Academic Press.

Lisker, Leigh. 1958. The Tamil occlusives: Short vs. long or voiced vs. voiceless? *Indian Linguistics, Turner Jubilee* I. 294–301.

Lisker, Leigh. 1986. "Voicing" in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29. 3–11.

Local, John and Adrian Simpson. 1999. Phonetic implementation of geminates in Malayalam nouns. The phonetic realization of singleton-geminate contrasts in three languages of Indonesia. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS 14), San Francisco: University of California, 595–598.

Löfqvist, Anders. 2006. Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *Journal of the Acoustical Society of America* 120(5). 2872–2883.

Löfqvist, Anders. 2007. Tongue movement kinematics in long and short Japanese consonants. *Journal of the Acoustical Society of America* 122(1). 512–518.

Maddieson, Ian. 1985. Phonetic cues to syllabification. In Victoria Fromkin (ed.), *Phonetic linguistics*, 203–221. London: Academic Press.

Maekawa, Kikuo. 1984. Akita hōgen sokuon no jizoku jikan: "Sunzumari" no jittai to seiin [Duration of geminates in the Akita dialect]. *Hōgen Kenkyū Nenpō* 27. 231–247.

Maekawa, Kikuo. 2010. Coarticulatory reintrepretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese. *Journal of Phonetics* 38(3). 360–374.

Martin, Andrew and Sharon Peperkamp. 2011. Speech perception and phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*. 2334–2356. Malden, MA and Oxford: Wiley-Blackwell.

Matsuura, Toshio. 2012. Yūsei sogai jūshiin no onsei jitsugen ni okeru chiikisa ni kansuru yobiteki bunseki [A preliminary analysis on regional variation in phonetic realization of voiced obstruent geminates]. Talk presented at the Annual Convention of the Phonetic Society of Japan, 29 September.

Mattei, Marco and Maria Gabriella Di Benedetto. 2000. Acoustic analysis of singleton and geminate nasals in Italian. *WEB-SLS: The European Journal of Language and Speech* (online).

Matsui, Michinao. 2012. Shakuyōgo ni okeru sokuon seiki no yokusei yōin [An inhibitory factor of occurrence of voiceless geminate consonants in Japanese loanwords]. *Theoretical and Applied Linguistics at Kobe Shoin* 15. 49–102.

Miller, Joanne L. and Alvin M. Liberman. 1979. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics* 25. 457–465.

Mori, Yoko. 2002. Lengthening of Japanese monomoraic nouns. *Journal of Phonetics* 30(4). 689–708.

Motohashi-Saigo, Miki and Debra M. Hardison. 2009. Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Training* 13(2). 29–47.

Muller, Jennier S. 2001. *The phonology and phonetics of word-initial geminates*. Columbus, Ohio: Ohio State University dissertation.

Nakajima, Yoshitaka, Gert ten Hoopen, and Gaston Hilkhuysen. 1992. Time-shrinking: A discontinuity in the perception of auditory temporal patterns. *Perception & Psychophysics* 51(5). 504–507.

Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Norlin, Kjell. 1987. A phonetic study of emphasis and vowels in Egyptian Arabic. *Lund University Department of Linguistics Working Papers 30*.

Oba, Reiko, Angelika Braun and Jürgen Handke. 2009. The perception of Japanese geminates by native and non-native listeners. *ISPhS: International Society of Phonetic Sciences*, 9–29.

Obrecht, Dean. 1965. Three experiments in the perception of geminate consonants in Arabic. *Language and Speech* 8. 31–41.

Ofuka, Etsuko. 2003. Sokuon /tt/ no chikaku: Akusento-gata to sokuon/hisokuongo no onkyōteki tokuchō ni yoru chigai [Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 7(1). 70–76.

Ofuka, Etsuko, Yoko Mori and Shigeru Kiritani. 2005. Sokuon no chikaku ni taisuru senkō kōzoku boinchō no eikyō [The effects of the duration of preceding and following vowels on the perception of geminates]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 9(2). 59–65.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer.

Ohala, John J. and Carol. J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. In Jared. J. Wolf and Dennis. H. Klatt (eds), *Speech communication papers*, 89–92. New York: The Acoustical Society of America.

Ohala, Manjari. 2007. Experimental methods in the study of Hindi geminate consonants. In Maria Josep Solé, Patrice Beddor and Manjari Ohala (eds.), *Experimental approaches to phonology*, 351–368. Oxford: Oxford University Press.

Otake, Takashi. this volume. Chapter 12: Mora and mora timing. In Haruo Kubozono, (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Otaki, Yoko. 2011. The boundary of the quantitative contrast between singleton/geminate consonants in production and perception. In Wai-Sum Lee and Eric Zee (eds.), *Proceedings of ICPhS XVII*, Hong Kong: City University of Hong Kong, 1546–1549.

Payne, Elinor. 2005. Phonetic variation in Italian consonant gemination. *Journal of International Phonetic Association* 35(2). 153–181.

Payne, Elinor. 2006. Non-durational indices in Italian geminate consonants. *Journal of International Phonetic Association* 36(1). 83–95.

Pickett, Emily, Sheila Blumstein, and Martha Burton. 1999. Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica* 56. 135–157.

Pickett, J. M. and Louis R. Decker. 1960. Time factors in the perception of a double consonant. *Language and Speech* 3. 11–17.

Pind, Joergen. 1986. The perception of quantity in Icelandic. *Phonetica* 43. 116–139.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Podesva, Robert. 2000. Constraints on geminates in Burmese and Selayarese. In Roger Bilerey-Mosier and Brook Danielle Lillehaugen (eds.), *Proceedings of West Coast Conference on Formal Linguistics 19*, 343–356. Somerville: Cascadilla Press.

Port, Robert and Johnathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception* & *Psychophysics* 32. 141–152.

Port, Robert and Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.

Port, Robert, Jonathan Dalby and Michael O'Dell. 1987. Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81. 1574–1585.

Ridouane, Rachid. 2010. Geminate at the junction of phonetics and phonology. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio and Nathalie Valleé (eds.), *Laboratory phonology 10*, 61–90. Berlin: Mouton de Gruyter.

Sagisaka, Yoshinori and Yoh'ichi Tohkura. 1984. Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyo [Phoneme duration control for speech synthesis by rule]. *Denshi Tsūshin Gakkai Ronbunshi* [The Transactions of the Institute of Electronics, Information and Communication Engineers A] 67(7). 629–636.

Sato, Hirokazu. 1978. Temporal characteristics of spoken words in Japanese. *Journal of the Acoustical Society of America [Abstract]* 64 (S1). 113–114.

Sawashima, Masayuki. 1968. Movements of the larynx in articulation of Japanese consonants. *Annual bulletin research institute of logopedics and phoniatrics* 31(1). 11–20.

Shirai, Setsuko. 2002. Gemination in loans from English to Japanese. MA thesis, University of Washington.

Shrotriya, Nisheeth, A. S. Siva Sarma, Rajesh Verma and S. S. Agrawal. 1995. Acoustic and perceptual characteristics of geminate Hindi stop consonants. In Kjell Elenius and Peter Branderud (eds.), *Proceedings of ICPhS XIII*, 132–135.

Smith, Caroline. 1995. Prosodic patterns in the coordination of vowel and consonant gestures. In Bruce Connell and Amalia Arvaniti (eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 205–222. Cambridge: Cambridge University Press.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30. 655–688.

Stevens, Kenneth and Sheila Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas and Joanne L. Miller (eds.), *Perspectives on the study of speech*, 1–38. New Jersey: Lawrence Erlbaum.

Stevens, Kenneth and Samuel Jay Keyser. 1989. Primary features and their enhancements in consonants. *Language* 65. 81–106.

Tajima, Keiichi, Hiroaki Kato, Amanda Rothwell, Reiko Akahane-Yamada and Kevin. G. Munhall. 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America* 123(1). 397–413.

Takada, Seiji. 1985. Sokuon no chōonjō no tokuchō ni tsuite [Articulatory characteristics of geminates]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kokuritsu Kokugo Kenkyūjo hōkoku 83: Kenkyūjo hōkokushyū 6*, 17–40. Tokyo: Kokuritsu Kokugo Kenkyūjo.

Takagi, Naoyuki and Virginia Mann. 1994. A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22. 343–356.

Takeyasu, Hajime. 2009. Masatsuon no sokuon chikaku ni okeru masatsu shūhasū tokusei no eikyō [Effects of the spectral properties of frication on perception of singleton/geminate fricatives]. *On'in Kenkyū* [Phonological Studies] 12. 43–50.

Takeyasu, Hajime. 2012. Sokuon no chikaku ni taisuru senkō onsetsu shiin/boin no jizoku jikan no eikyō [Effects of the consonant and vowel durations in the preceding syllable on the perception of geminate stops in Japanese]. *On'in Kenkyū* [Phonological Studies] 15. 67–78.

Tews, Andrea. 2008. Japanese geminate perception in nonsense words involving German [f] and [x]. *Gengo Kenkyu* 133. 133–145.

Tserdanelis, Georgios and Amalia Arvaniti. 2001. The acoustic characteristics of geminate consonants in Cypriot Greek. In Aggouraki, Yoryia, Amalia Arvaniti, J. I. M. Davy, Dyonysis Goutsos, Marilena Karyolaimou, Anna Panagiotou, Andreas Papapavlou, Pavlos Pavlou and Anna Roussou (eds.), *Proceedings of the Fourth International Conference on Greek Linguistics*, 29–36. Thessaloniki: University Studio Press.

van Dommelen, Wim. 1999. Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *Journal of Phonetics* 27. 107–123.

Warner, Natasha and Takayuki Arai. 1999. Japanese mora-timing: A review. *Phonetica* 58. 1–25.

Watanabe, Shinichiro and Nobuo Hirato. 1985. The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and the duration of the preceding vowels. *Onsei Gengo* 1. 1–8.

Westbury, John R. and Patricia Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22. 145–166.

Whalen, Douglas, Arthur Abramson, Leigh Lisker and Maria Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America* 47. 36–49.

Yu, Alan. 2011. Contrast reduction. In John Goldsmith, Jason Riggle and Alan Yu (eds.), *The handbook of phonological theory*, 2nd edition, 291–318. Oxford: Blackwell.

Itsue Kawagoe

2 The phonology of *sokuon*, or geminate obstruents

1 Introduction

Long consonants or geminate obstruents are abundant in Modern Japanese, as in /suppai/ 'sour', /kitte/ 'stamp' and /gakki/ 'music instrument'. However, in Old Japanese it is generally assumed that there were no geminate obstruents. Linguists tend to agree that Old Japanese had simple syllable structures, open syllables of the form (C) V, with the onset C as an optional segment (Hashimoto 1950; Yamaguchi 1989; among others) and there were no coda consonants. However, during the four centuries of Middle Japanese between the ninth and twelfth centuries, two types of coda consonants appeared, the moraic obstruents and the moraic nasals, introducing closed syllables into Japanese phonology (see Takayama, this volume, for the history of Japanese phonology). Moraic obstruents are often called *sokuon* in the traditional Japanese phonology. They are usually combined with the onset consonant of the following syllable to form long consonants phonetically and geminates phonologically. In this chapter, we will use the term coda obstruents or geminate obstruents.

It is widely agreed that both types of coda consonants appeared as a result of two phonological changes which took place in Middle Japanese, resulting in the contraction of the final syllables of verb and adjective stems (see Komatsu 1981; Yanagida 2003; among others). One changed CV syllables to either /i/ or /u/, thus reducing a CVCV sequence to a CVV sequence, as /kiki-te/ → /kii-te/ 'hear' (where /te/ denotes a gerundive ending and '-' denotes a morpheme boundary). The other changed stem-final CV syllables to coda consonants, thus producing a CVC syllable ending with a geminate or coda nasal, as /tori-te/ → /tot-te/ 'take' or /tobi-te/ → /ton-de/ 'fly', respectively.

The introduction of CVC syllables into Japanese was certainly influenced by a massive influx of Chinese vocabulary at that time, with its many CVC syllables, but recently language-internal motivations for coda consonants have also been pointed out. Komatsu (1981: 173) argues that both types of coda consonants existed in the Japanese mimetic vocabulary at the time of the phonological changes. Incoming Chinese words caused syllables with coda consonants to become more salient, and they were eventually incorporated into Japanese phonology as legitimate syllables. Hizume (2003: 101) suggests another factor, claiming that the closure duration of stop consonants was prolonged to give more emphasis to mimetic expressions,

which further encouraged gemination to become a phonological change. In this view, coda obstruents, or geminates, already existed before the phonological changes in the Japanese mimetic vocabulary as phonetically long consonants and they spread to the full native vocabulary.

Not all obstruents in Japanese can appear in the coda position. In the native and Sino-Japanese (henceforth SJ) vocabularies, only voiceless obstruents are allowed as geminates: [p], [t], [k], [s], [ʃ], [ts], and [ʧ]. In foreign loanwords, voiced obstruents, [b], [d], [g], [dz], and [ʤ], also can appear in this position[1] (see sections 6.2.2 and 6.4 for this issue).

This chapter provides an overview of the distribution of coda obstruents in each of the four lexical strata in contemporary Japanese, with particular focus on the loanword stratum, where gemination is most commonly observed. We will address why coda obstruents appear in certain positions in each lexical stratum, and discuss their functions therein.

To achieve this goal, the chapter is organized as follows. Section 2 will look at the phonological properties of coda obstruents, and review two phonological analyses. Section 3 will discuss coda obstruents in the native vocabulary, including verb inflection and compounds. We will see that coda obstruents serve to maintain prosodic structure and help define the integrity of compounds. Section 4 will discuss coda obstruents in SJ compounds with the introduction of the concept of contraction under feature compatibility. SJ compounding will be contrasted with a similar process in the native vocabulary, revealing the different nature of these two processes. Section 5 will discuss coda obstruents in the mimetic vocabulary with focus on their peculiar behaviors. Section 6 will be concerned with the distribution of coda obstruents in loanwords, outlining two relevant analyses of their occurrence.

2 Phonological status of geminates

Traditional Japanese phonologists have analyzed the coda obstruents as one phoneme called *sokuon* and represented it with /Q/[2] (Arisaka 1940; Hattori 1960; Hashimoto 1950; Koizumi 1978; among others). Thus, for example, /kitte/ 'stamp' is represented as /kiQte/, and /gakki/ 'emergency' as /gaQki/. Evidence for the phonemic status of the coda obstruent comes from such minimal pairs as listed in (1).

1 Marginally, [h] appears in coda position, as in /zjuhhari/ 'ten stitches' in the native vocabulary and /gohho/ 'van Gogh' in loanwords.

2 In this chapter the symbol /Q/ is used to represent the first half of the geminate consonants when referring to the representation in the traditional phonology. Besides that, /CC/ is used.

(1)	a.	[p/h]	[ippai] 'one defeat'	[ihai] 'a mortuary tablet'[3]
	b.	[t]	[ittai] 'a party'	[itai] 'a corpse'
	c.	[k]	[ikkai] 'the first floor'	[ikai] 'underworld'
	d.	[ʧ]	[itʧi] 'agreement'	[iʧi] 'one'
	e.	[s]	[issai] 'everything'	[isai] 'details'
	f.	[ʃ]	[iʃʃi] 'one child'	[iʃi] 'volition'

Geminate and single consonants in each pair in (1) are in parallel (overlapping) distribution and serve to signal a semantic contrast. Thus the pairs in (1) make minimal pairs, showing that the first halves of the geminates, i.e., the coda obstruents are phonemes. Also, the coda obstruents in the left-column words, do not contrast with each other, so they are in complementary distribution, showing that they are different realizations of one same phoneme. To see the complementary distribution clearly, let us look at the coda segment [p] in [ippai] (1a). It appears only before onset [p] and never before onset [t], as in [ittai] (1b), *[iptai]. The same is true of the coda segment [t] in [ittai] (1b): it never appears before the onset segment [p], *[itpai]. Coda segments [p] and [t] are complementarily distributed and so are all the coda consonants in (1). Since consonant length is contrastive and coda obstruents are in complementary distribution with each other, all the coda obstruents are grouped together into one phoneme /Q/, and each coda obstruent is called an allophone of /Q/.

As Vance (2008: 107) writes: "/Q/ seems like a chameleon phoneme to an English speaker because it has such a wide variety of phonetically different realizations." When /Q/ is followed by a plosive, it shows "an abrupt suspension of articulatory movements" (Arisaka 1940: 94), and when it is followed by a fricative, air continues to flow from the lungs (Vance 1987: 40). Admitting the difference among the allophones of /Q/, it is generally agreed that "they share enough phonetic similarity to make it plausible to treat them all as realizations of the same abstract entity" (Vance 2008: 106), which is an unreleased long obstruent. Accepting that these qualities are shared by the allophones of /Q/, still it is peculiar that the phoneme /Q/ does not have any phonological specifications other than being consonantal and non-nasal. Koizumi (1978: 114) notes that /Q/ is quite peculiar in that it does not have its own phonetic substance and that it is a fabricated phoneme to integrate its allomorphs.

Phoneme /Q/ is peculiar not only in its phonetic substance, but also in its restricted occurrence in a syllable: it always occurs in the coda position, and never in the syllable-initial position, and is always followed by another consonant.[4] Because of this positional restriction, /Q/ is called a special or dependent phoneme different from regular phonemes. Coda nasals and the second parts of long vowels/diphthongs are also treated in the same way. /Q/, like these dependent phonemes,

3 Geminated /h/ is normally replaced by its morphophonemic alternant /pp/, not by /hh/, but for some special cases, /hh/ appears, as exemplified in note 1.

4 In some special cases, /Q/ can appear utterance-finally as a glottal stop, but as Labrune (2012: 135) argues, its function in these cases is expressive and not distinctive.

counts as one mora in length just as a CV mora does; hence, they are also called "moraic phonemes" (Jōo 1977: 120).

Although the analysis of geminates using phoneme /Q/ does not provide any theoretical explanation of their peculiarities, the idea of phoneme /Q/ seems to be deeply rooted in the native speaker's intuition. This is clearly noted in Vance (2008: 107): "Native speakers of Japanese feel that a syllable-final obstruent… isn't the same as any syllable-initial consonant.… The phoneme /Q/ is a straightforward reflection of these intuitions." As Vance (1987: 41) states, kana orthography might be supporting the intuition of native speakers, since the allophones of /Q/ are all written the same, using the letter for the syllable *tsu* reduced in size.

A geminate consonant consists of two parts: the first part is called "sokuon" in the traditional analyses (symbolized as /Q/) and "coda obstruent" in later syllable-based analyses, while the second part is a regular obstruent constituting the onset of the following syllable. There are three special properties of geminates that must be accounted for, all concerning the characterization of the first part. The traditional analyses meet these requirements by positing a special phoneme /Q/ with moraic status and no phonetic content, to be followed by another obstruent.

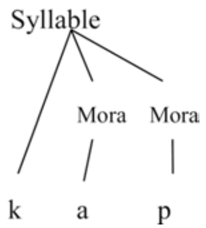
(2) The coda obstruent in a geminate
  a. is an oral obstruent with no place of articulation of its own;
  b. is always followed by another obstruent;
  c. has a moraic status, occupying one mora.

Now we consider modern characterizations of geminates, which offer a more satisfying explanation of the three special properties. In general, generative phonology accounts for allomorphic relations by establishing one and the same underlying form and deriving surface forms by rules. McCawley (1968) and Kuroda (1965) argued in their morphophonemic analyses of Japanese phonology that special phonemes like /Q/ are not represented as such at any level of phonological derivation. Optimality Theory (Prince and Smolensky 1993, hereafter "OT") in general extends the principles of derivation-based generative phonology and uses no special phonemes in its analysis of geminates. Thus, the geminated forms in (1a–f) are shown on the left-hand side of (3) as underlying morphemes, with their corresponding surface forms on the right. All six of these SJ compounds have the same initial morpheme /it/ 'one' (hyphens indicate morpheme boundaries), and the surface forms are derived by an assimilation process which we will see in section 3.1. In this theory, there is no level for a phoneme /Q/ to be represented.

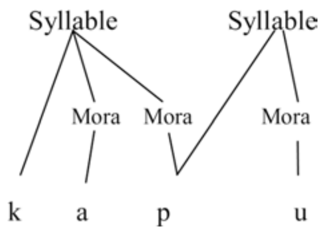
(3)	a.	/it-hai/	[ippai]	'one defeat'
	b.	/it-tai/	[ittai]	'one party'
	c.	/it-kai/	[ikkai]	'the first floor'
	d.	/it-ti/	[itʧi]	'agreement'
	e.	/it-sai/	[issai]	'everything'
	f.	/it-si/	[iʃʃi]	'one child'

OT is based on the moraic hypothesis (Hyman 1985; McCarthy and Prince 1986; and others) which assumes that a syllable internal structure involves the mora. Light syllables consist of one mora, whereas heavy syllables consist of two moras. (4a) shows a representation of a syllable with segments linked to moras. The onset consonant, i.e., /k/ in (4a), is directly linked to the syllable node with no linking to the mora node. The segment in the coda position, i.e., /p/ in (4a), is universally more restricted than the one in the onset position. Ito and Mester (1993) propose conditions for licensing coda consonants in Japanese: A segment in the coda position is licensed so long as (i) it is a nasal or a vowel and (ii) it has no independent consonantal place feature. However, if such a consonant is doubly linked and one of the two links is licensed, as is a link to the following onset, then the segment is well-formed. Thus, the form */kap/ in (4a) is ill-formed in Japanese because (i) the coda consonant /p/ is not a nasal, and (ii) it has its consonantal place feature.

(4) a. */kap/



b. kappu 'cup'



However, the form /kappu/ in (4b) is well-formed because the consonant /p/ here is doubly linked, i.e., not only to the coda position of the first syllable but also to the onset position of the second syllable. The linking of the segment /p/ to the coda position is not licensed, but the one to the onset position is licensed. Thus the segment /p/, although it has a place feature, is well-formed in (4b).

This account explains the three special properties of the first part of the Japanese geminate shown in (2). In Japanese, coda obstruents are well-formed only when they are doubly linked, that is, linked not only to the coda position but also to the onset position of the following syllable. Without double linking, a consonant with a place feature in coda position is ill-formed. This explains why a coda obstruent can have no place of articulation of its own, but must always be followed by another obstruent, and does not appear word-finally except in special cases (see note 4). Moreover, the

moraic status of the coda obstruent derives from the basic assumptions of moraic theory. For all three properties of coda obstruents, this analysis uses coda licensing conditions that restrict segments in this particular position.

Although the coda consonant has moraic status in modern Tokyo Japanese, as we have seen, this has not always been the case historically, and is not even true for all modern dialects of the language. As shown in section 1 above, Japanese syllables originally had a very simple CV form. When CV was a canonical syllable and at the same time a mora, CV-based timing was easily established (Yamaguchi 1989: 1647). With the emergence of coda consonants, there were two possible ways to organize the temporal structure of speech. One was to assign two moras to the new CVC syllables, with the coda consonant counted as an independent mora. The other was to interpret both CV and CVC similarly as a single timing unit, yielding a system called syllable-based timing. It is asserted (Okumura 1977: 231) that when coda consonants were introduced, the single-unit system was in practice, but with the gradual establishment of the phonemic status of coda consonants, the dual-unit system took over. In Modern Japanese some dialects, such as the one in Kagoshima, still keep the single-unit system, where both CV and CVC syllables are treated as one timing unit (Kubozono 1999: 153). In the rest of this chapter, we will follow the OT analysis of coda obstruents and treat them as geminates, with no further reference to the phoneme /Q/.

3 Geminates in the native vocabulary

The basic form of the syllable in Japanese is an open syllable, i.e., a syllable that ends in a vowel. Closed syllables are infrequent in native morphemes, which means that native monomorphemic words with geminates are quite rare. In the native vocabulary, geminates occur most typically in suffixed forms and in compounds, as well as in intensified forms of adverbs and mimetics.

3.1 Geminates in verb inflection

Verbs in the native vocabulary are classified into two types, vowel-final and consonant-final verbs (cf. Bloch 1946: 7).[5] The stems of the former end in vowels, like /mi/ 'look at' and /tabe/ 'eat', and those of the latter end in consonants, like /kir/ 'cut'. Stem-final consonants are restricted to the following nine: /w/, /r/, /m/, /n/, /s/, /t/, /k/, /b/, /g/.

5 Japanese has very few irregular verbs, two of which, /kuru/ 'come' and /suru/ 'do', are the most often cited. According to Tamamura (1989: 43), 5.38% of the 3457 verbs that appeared in the ninety magazines of 1960s was irregular. In this section we will deal only with regular verbs.

Let us consider what happens when a suffix of the form CV attaches to a verb stem. When the past-tense suffix /ta/, for example, attaches to the vowel-final verb stem /mi/, the underlying /mi-ta/ surfaces as [mita], with no morphophonological changes (hyphens mark the morphological boundary between the stem and the suffix). However, when /ta/ attaches to the consonant-final verb stem /kir/ 'cut', underlying /kir-ta/ surfaces as [kitta], with the underlying /rt/ sequence realized as a geminate.[6] In verb stems ending with the nine possible stem-final consonants, the three types of phonological changes in (5) are observed. In (5) and the rest of the chapter, dots indicate syllable boundaries.

(5)	a.	kir-ta	'cut'	→	kit.ta
		kaw-ta	'bought'	→	kat.ta
		kat-ta	'won'	→	kat.ta
	b.	sin-ta	'died'	→	sin.da
		yob-ta	'called'	→	yon.da
		yom-ta	'read'	→	yon.da
	c.	kak-ta	'wrote'	→	kai.ta
		kag-ta	'sniffed'	→	kai.da
		kas-ta	'lent'	→	ka.si.ta

In (5a), where the stems end in /r/, /w/ or /t/, gemination occurs, while in (5b) stems end in /n/, /m/ or /b/ and a coda nasal appears. In (5c), /i/ is epenthesized when the stem ends in /k/, /g/ or /s/, and, in addition, the velar coda consonant /k/ or /g/ deletes before the epenthetic vowel /i/; instead of *[kakita] or *[kagita], [kaita] and [kaida] surface, with further alternation of /t/ to /d/ in the latter case. On the other hand, /s/ in (5c) does not delete, so [kaʃita] surfaces, with /s/ phonetically realized as [ʃ]. Voicing of the suffix takes place in (5b) and also after the voiced velar consonant (/kag-ta/→/kai.da/) in (5c).

Let us now consider why gemination occurs in (5a), but not in (5b) and (5c). When the past-tense suffix /ta/ is attached to a consonant-final verb stem, two consonants are juxtaposed: /rt/, /wt/, /tt/, /nt/, /bt/, /mt/, /kt/, /gt/, or /st/. Except for the /tt/ form, which surfaces as a geminate, all these forms are phonotactically illegal in Japanese. To fix the impermissible CC sequence, two processes are observed: (i) regressive assimilation of the two CC consonants, deriving either a coda obstruent or a coda nasal, or (ii) insertion of a vowel between the two consonants, deriving a CVC sequence. The forms in (5a) and (5b) undergo the assimilation process, while those in (5c) undergo the epenthesis process. Why do the forms in (5a) and (5b) not undergo epenthesis, while those in (5c) do?

6 Morphophonological changes are triggered by a suffix beginning with /t/. Some other consonant-initial verbal suffixes are *te* (gerundive suffix), and *tari* (alternative suffix). They are related to the past suffix /ta/ (Vance 1987: 184).

Two attempts to answer this question have been put forward, neither of which seems completely satisfactory. One approach attempts to find out why vowel epenthesis takes place only in the three sequences in (5c). This approach, pursued by McCawley (1968: 96) and Aoki (1981: 71), looks for phonological properties that would group these three stem types into a natural class and explain why they go together through vowel epenthesis. According to McCawley, underlying segments /k/ and /g/ are first converted to /h/, which has feature [+continuant], and this is also a feature of /s/ and is thus shared by all three stem types that undergo epenthesis. Aoki (1981: 67) argues that there is "no synchronic or diachronic evidence given to support such claims," and proposes that the feature [-anterior] is the one shared by /k/, /g/, and /s/. Aoki claims that Japanese /s/ is [-anterior], but Tsujimura and Davis (1988: 490) criticize this hypothesis as lacking any independent evidence.

Both McCawley and Aoki assume that assimilation produces coda consonants unless blocked by epenthesis. On the other hand, the second approach, which is taken by Davis and Tsujimura (1991: 118), posits that epenthesis occurs unless blocked by assimilation and thus asks why assimilation takes place for sequences /rt/ and /wt/ as in (5a) and /mt/, /bt/ and /nt/ in (5b). Under their analysis, in (5a), non-nasal sonorants /r/ and /w/ are delinked from the C-slot on the CV-tier before /t/ (see Davis 2011 for detailed discussion of the CV-tier), leading to total assimilation, while in (5b), labials /m/ and /b/ change to dentals and then to [n], before [t]. Since the place of articulation features of [n], [+coronal, +anterior] are shared by the following [t], assimilation takes place and epenthesis is blocked. On the other hand, the forms in (5c) with /kt/, /gt/ and /st/ do not share the same place of articulation, cannot be assimilated, and so are epenthesized. Note that /t/ in Japanese is dental and differs from /s/, which is alveolar.

Neither of these explanations offers independent evidence to support their arguments. Why is it a specific group of segments that undergoes vowel epenthesis and not others? Why are non-nasal sonorants /r/ and /w/ delinked from the C-position (sonorant delinking), and why do labial obstruents (/b/) undergo nasalization (nasal linking)? These questions have not been addressed to a satisfactory degree. What is clear is that gemination in verb inflection is created by assimilation, one of the two processes which serve to change an impermissible CC sequence created by suffixation to a phonotactically legal structure.

3.2 Geminates in verb-verb compounds

We have described the various phonological changes observed in verb inflection when two consonants are concatenated and the second one is /t/. In this section we will see what happens when two consonants are concatenated due to verb-verb compounding in the native vocabulary.

The verb-verb compounds illustrated in (6a–c) have nine verb stems with CC across the morpheme boundary and with /t/ as the second one, just as in (5a–c). Here, the first verb stem must end in a consonant i.e., a consonant-final verb, in order to make a CC sequence (see Poser 1984: 89 and Vance 2002: 367 for details). In (6a), the first consonant is either /t/ or the non-nasal sonorant /r/ or /w/; in (6b), it is either a nasal or a labial; in (6c), it is either /k/, /g/, or /s/. These are the same conditions presented above with examples of verb inflection. The final /u/ in (6) and below is a non-past suffix.

(6) Verb-verb compounds
   a. kir-tor-u → kiri-toru 'to cut off'
      kaw-tas-u → kai-tasu 'to make additional purchases'
      kat-tor-u → kati-toru 'to win'
   b. sin-taer-u → sini-taeru 'to die out'
      yob-tater-u → yobi-tateru 'to summon'
      yom-tor-u → yomi-toru 'to read between the lines'
   c. kak-tas-u → kaki-tasu 'to add a few more lines'
      kag-tor-u → kagi-toru 'to sense'
      kas-tuker-u → kasi-tukeru 'to loan'

As can be seen from these examples, all the CC sequences undergo /i/ epenthesis, surfacing as CiC sequences irrespective of the phonological properties of the first consonant C, in contrast to the results in verb inflection we saw above. /i/ epenthesis also occurs in verb-verb compounds with a CC sequence where the second consonant is other than /t/. To take the verb stem /kir/, for example, epenthesis applies no matter what the following consonant is: /kir-hanas-u/ → /ki.ri.ha.na.su/ 'to cut off'; /kir-naos-u/ → /ki.ri.na.o.su/ 'to cut all over again'; /kir-kaker/ → /ki.ri.ka.ke.ru/ 'to begin to cut, to be about to cut'; /kir-sak-u/ → /ki.ri.sa.ku/ 'to cut up'. The same is true of the other verb stems in (6) and indeed in most verb-verb compounds. This suggests that vowel epenthesis is the default strategy to fix up the impermissible sequence of CC for verb-verb compounding.

However, one type of native verb-verb compound does surface with geminates. Vance (2002: 368) notes that geminates surface only when the first stem is two moras long and only for a small number of stems, two of which are illustrated in (7). These are called verbal root compounds in Ito and Mester (1996: 24). The left-hand verbal root usually adds the meaning of 'intense action' and colloquial flavor to the meaning of the right-hand root.[7] / / denotes underlying representations rather than phonemic ones here.

7 Some compounds with geminates but with no emphatic meaning are /not-tor-u/ 'take over', /hik-kos-u/ 'move into', /sap-pik-u/ 'subtract'. Vance (personal communication).

(7) Verbal root compounds (Ito and Mester 1996: 24)

a.	/but-/ 'strike'	/taos-u/ 'let go down'	but-taosu	'knock down'
		/koros-u/ 'kill'	buk-korosu	'kill violently'
		/nagur-u/ 'beat'	bun-naguru	'beat forcefully'
b.	/hik-/ 'pull'	/tuk-u/ 'stick to'	hit-tuku	'stick close together'
		/kak-u/ 'scratch'	hik-kaku	'scratch violently'
		/sak-u/ 'tear'	his-saku	'tear apart forcefully'
		/muk-u/ 'peel'	him-muku	'peel off violently'

In these compounds, assimilation takes place when two consonants are juxtaposed to form an impermissible consonant cluster, hence yielding a geminate.[8] One noteworthy aspect of this assimilation is that the juxtaposed consonants are not limited to a particular set as in verb inflection in (5), but all combinations of consonants result in gemination. Thus, if the second consonant is a sonorant, the first obstruent changes its sonority, producing a geminate sonorant, as seen in /but-nagur/ → /bunnaguru/ (7a). As we will see below, this does not happen for assimilation in the SJ vocabulary: e.g., /bet-noo/ → /betunoo/, */bennoo/ 'spreading payment'. This difference will be discussed in section 4.3 below.

3.3 Geminates in other compounds

Not only verb-verb compounds but also noun and other compounds show gemination, as exemplified in (8a) (Takayama 1995: 17):

(8)	a.	hitori	'one person'	ko	'child'	→	hitorikko	'an only child'
		kimo	'courage'	tama	'ball'	→	kimottama	'courage'
		mie	'vanity'	hari	'stretch'	→	mieppari	'vain'
		huki	'blow'	sarasi	'exposure'	→	hukissarasi	'windswept'
	b.	sute	'abandon'	ko	'child'	→	sutego	'an abandoned child'
		ame	'candy'	tama	'ball'	→	amedama	'ball-shaped candy'
		sita	'beneath'	hari	'stretch'	→	sitabari	'base sheet'
		ame	'rain'	sarasi	'exposure'	→	amazarasi	'weather-beaten'

8 Occasionally, a particular combination of verbal roots results in two different compounds, one with a geminate and the other without. The forms with geminates usually have the more intensified meaning. In some cases, a semantic difference other than emphasis in meaning is observed, as seen in the pairs below (Saito 1992: 227). See also Vance (2002: 373) for more semantic divergences of the verb-verb compound pairs.

/hik-/ 'pull' + /tuker-u / 'attach'	→	hit-tukeru	'glue together'
	→	hiki-tukeru	'attract'
/huk-/ 'blow' + /tob-u / 'fly'	→	hut-tobu	'vanish'
	→	huki-tobu	'blow off'

Gemination occurs in forms like (8a) only if the initial segment of the second element is a voiceless obstruent. It never occurs if the subsequent element starts with a vowel or a nasal (e.g., /hitori-ne/, */hitorinne/ 'solitary sleeping'). The compounds in (8b) have the same second elements as those in (8a), but show no gemination; rather, we see voicing of the initial consonant, called *rendaku*, or sequential voicing (see Vance, this volume, for more details about this mophophonological process). Since Japanese phonotactics prohibits word-initial voiced obstruents in the native vocabulary, *rendaku* in these compounds shows that the second element is no longer word-initial, thus signaling the merger of the two elements into one compound. Likewise, geminates, which can never appear in word-initial position, indicate a compound-internal boundary. Gemination and sequential voicing are claimed to play the same role in compounding (Komatsu 1981: 174; Takayama 1995: 17).

4 Geminates in Sino-Japanese compounds

During its long history of borrowing from the Chinese language, Japanese accumulated a large number of SJ compounds. These compounds are created by combining SJ word stems (or roots).[9] The stems are of the form $(C_1)V_1C_2(V_2)$, which have a certain morphemic structure and morphophonemic characteristics (Ito and Mester 1996: 14): (i) a prosodic size limit, i.e., maximally two moras, (ii) predictable V_2, and (iii) a neutralization of consonant features in C_2 position. (iii) is a process of contraction applying almost without exception in this lexical domain and yielding geminates. In this section, we present the basic facts of contraction (gemination), followed by its phonological analysis, which will be contrasted with a similar process in the native vocabulary. Some problems of this analysis will be discussed at the end of this section.

4.1 Basic facts about SJ contraction

In SJ stems, the final vowel is mostly predictable and thus assumed to be epenthetic (Ito 1986; Tateishi 1990). In contemporary Japanese, only /t/ and /k/ can appear as the second consonant of SJ stems, classified as t-stems and k-stems, respectively, and contraction takes place for both by the rules of Japanese phonology. In t-stems, contraction takes place with any following voiceless obstruent, but not with any other segment type. In k-stems, contraction takes place only when the k-stem is followed by another /k/.

9 According to Vance (personal communication), the term 'root' is a better choice here, since traditionally a stem is a base for inflectional forms. But we use 'stem' in accordance with Ito and Mester (1996).

Consider the t-stem forms in (9), all with /hat-/ as their initial stem. Only those whose stem is followed by a voiceless obstruent as in (9a) result in geminates; those in (9b) do not show gemination.

(9) /hat(u)/ 'start'

a.	/hat-poo/	happoo	'open fire'
	/hat-tat/	hattatu	'development'
	/hat-ken/	hakken	'discovery'
	/hat-hun/	happun	'be stimulated'
	/hat-san/	hassan	'emission'
b.	/hat-bai/	hatubai	'putting on sale'
	/hat-den/	hatuden	'generation'
	/hat-gen/	hatugen	'remark'
	/hat-zyoo/	hatuzyoo	'sexual excitement'
	/hat-mei/	hatumei	'invention'
	/hat-rei/	haturei	'issue an order'
	/hat-an/	hatuan	'suggestion'
	/hat-wa/	hatuwa	'utterance'

In (10) all the examples shown are k-stems, all /rak-/, and contraction (gemination) takes place only when the stem is followed by /k/ as in (10a). When the stem is followed by any other segment, there is no contraction as shown in (10b).

(10) /rak(u)/ 'falling'

a.	/rak-ka/	rakka	'falling'
	/rak-kei/	rakkei	'celebration of the completion of a temple'
b.	/rak-hak/	rakuhaku	'reduction in poverty'
	/rak-tan/	rakutan	'disappointment'
	/rak-seki/	rakuseki	'rock fall'
	/rak-ba/	rakuba	'falling off one's horse'
	/rak-dai/	rakudai	'failing an examination'
	/rak-go/	rakugo	'dropping out'
	/rak-za/	rakuza	'free markets'
	/rak-mei/	rakumei	'death'
	/rak-rai/	rakurai	'being struck by lightning'
	/rak-yoo/	rakuyoo	'fallen leaves'
	/rak-in/	rakuin	'an illegitimately born child of royal blood'

4.2 Contraction under feature compatibility

The issue of SJ contraction was originally discussed in McCawley (1968), and then studied extensively, both in rule-based phonology (Ito 1986; Cho 1989; Ito and Mester

1996; and others) and in constraint-based OT (Nasu 1996; Makihara 1998; Kurisu 2000). The contraction results listed in (9) and (10) look like two different phenomena: those in (10) take place only when the stem-final /k/ is followed by /k/, while those in (9) take place when the stem-final /t/ is followed by any voiceless obstruent. Thus Ito (1986), Tateishi (1990) and Cho (1989) analyze them as two distinct processes. However, Ito and Mester (1996: 23) argue that the two contraction behaviors should be attributed to the same mechanism or process called "root fusion" by which the operational differences follow from the different representations of /t/ and /k/.

A quick glance at the k-stem contractions in (10a) shows that the coda /k/ fuses with the following /k/ under feature identity. The same is true with /t-t/, as in /hat-tatu/ (9a). However, in cases such as /hat-poo/, /hat-ken/ and /hat-san/ in (9a), the consonant pairs /t-p/, /t-k/, and /t-s/ are not identical. Contraction in these cases takes place not under feature identity, but under feature compatibility, which assumes underspecification of /t/ in term of place features, i.e., /t/ and /k/ are feature compatible because stem-final /t/ has no place specifications.

As mentioned above, in the SJ vocabulary, the stem-final obstruent is restricted to either /k/ or /t/. With cross-linguistic evidence for [coronal] as the default place feature (Paradis and Prunet 1991), and taking into account the fact that stem-final /t/ triggers contraction with any following voiceless obstruent, Ito and Mester (1996: 31) assume that stem-final /t/ has no place specification and is given the feature [coronal] by a universal default, while stem-initial /k/ is specified as [dorsal]. With this underspecification, the fusion of /t-p/, /t-k/ and /t-s/ are claimed to take place under feature compatibility. Thus, in the form /hat-ken/, the [dorsal] feature of consonant /k/ is compatible with the preceding underspecified /t/ and fuses with it.

Note that while stem-final /t/ is underspecified for place, stem-initial /t/ is not. In stem-initial position, full specification of place and manner features is found. In particular, place features are fully contrastive in this position. Thus, root fusion takes place in the form /hat-ken/ but not in the form /hak-tai/ → /hakutai/ 'fur on the tongue', because stem-initial /t/ has the place feature [coronal] and the two segments /k/ and /t/ are thus not feature-compatible (Ito and Mester 1996: 29–31). The fusion of /t-t/ and /k-k/, where the two consonants are identical, can also be analyzed as a case of fusion under feature compatibility.

By introducing the concept of feature compatibility and the underspecification of place features of stem-final /t/, the two seemingly distinct processes of contraction in /t/-stem and /k/-stem forms can be accounted for as one and the same phenomenon.

4.3 Root fusion vs. root spreading

In section 3.2, we discussed gemination in native verbal root compounds like /but-koros/ → /bukkorosu/ in (7a), a result which looks quite similar to that in the SJ

compounds in (9) and (10). However, there are two differences between these two cases. First, as pointed out by Ito and Mester (1996: 24), gemination in native root compounding takes place even when the second consonant is a sonorant, resulting in geminate sonorants like /but-nagur/ → /bunnaguru/ (7a). This does not happen in SJ compounds: /but-noo/ → /butunoo/, */bunnoo/ 'spreading payment'; /hat-mei/ → /hatumei/, */hammei/ in (9b). The other difference concerns stem-final /k/, which triggers gemination of any following consonant in native root compounding, e.g., /hik-tuk/ → /hittuku/ in (7b), but assimilates only with a following /k/ in SJ compounding: /rak-tan/ → /rakutan/, */rattan/ in (10b).

According to Ito and Mester (1996: 24–30), these differences arise because they involve different contraction processes: root fusion in SJ compounding and root spreading in native verbal root compounding. Root fusion, where two segments fuse into one, takes place under feature compatibility, as we saw above. When the features of the relevant two segments are not compatible, root fusion does not take place, like /t/ [-sonorant] followed by /n/ [+sonorant], or the case of /k/ [+dorsal] followed by /t/ [+coronal], as we have seen above in the case for SJ compounds.

Root spreading, on the other hand, occurs in native verbal root compounding and is initiated by a complete delinking of the root of the stem-final segment. It leads to complete assimilation even when sonority is different between the two segments or when the stem-final segment is /k/ followed by any [+consonantal] segment. In root spreading, the consonantal features of the onset spread regressively to obliterate those of the stem-final segment, so that the coda segment's characteristics are not easy to recover. In SJ root fusion, by contrast, stem-final obstruents /k/ and /t/ are always recoverable because they never contract with sonorants.

In short, it is claimed that (i) contraction (gemination) of t-stems and k-stems in SJ compounding is one and the same process, but (ii) the gemination found in native verbal root compounding is a different process.

4.4 Some problems

The analysis by Ito and Mester (1996) is a radical one in the history of the research of contraction in SJ compounding. Traditionally, the stem-final /i/ and /u/ were regarded as belonging to the underlying lexical forms of the SJ morphemes and vowel deletion rather than epenthesis was assumed to occur in the SJ compounding. Ito and Mester (1996) argued against this analysis and claimed instead that the stem-final /i/ and /u/ are almost entirely predictable: It is always /u/ except when the stem final /k/ is preceded by the front vowel /e/ like /teki/ 'an enemy' and /eki/ 'fluid'. They thus claimed that a simple analysis can be maintained by considering the stem-final vowels as epenthetic.

Kurisu (2000: 154) and Labrune (2012: 31), building on Vance (1987: 160), cast doubt on the predictability of the final vowels. They pointed out that forms like

/niti/ 'sun' and /hati/ 'eight' have final /i/ rather than the predicted /u/ and, moreover, that there are forms with /u/~/i/ alternations like /kiti/~/kitu/ 'good fortune'. Labrune (2012: 31) also argues that some of these final vowels carry accent, which epenthetic vowels usually don't do. Based on these pieces of evidence, they argue that final vowels of SJ stems are not predictable and thus not epenthetic, but exist in the lexical forms.

Another problem is the numerous exceptions to contraction in SJ compounding. The exceptions are of two types: the forms which regularly refuse to contract and those where contraction is inconsistent. In general, as Vance (1987: 158–159) notes, contraction is very nearly automatic when the first morpheme has the final vowel /u/, but many exceptions are observed when the first morpheme has the final vowel /i/. This is true with the first type of exceptions, e.g., /eki-ka/ 'liquification' not */ek-ka/, /teki-kaku /'qualified' not */tek-kaku/, and /eki-kin /'profits' not */ek-kin/. Their first morphemes end in /i/ and they refuse to contract. However, the second type of exceptions include not only words whose first morphemes end in /i/ like /tek-ki/~/teki-ki /'enemy flag', /gek-ka/~/geki-ka/ 'intensification', but also words whose first morphemes end in /u/ like /kak-kai/~/kaku-kai/ 'the sumo world', /jak-ka/~/jaku-ka/ 'prices for medicines' and /kak-ko/~/kaku-ko/ 'each home'. The final vowel of the first morpheme signals the exceptionality of the morpheme to contraction in many cases, but it has its limitations.

One interesting fact about the occurrence of contraction is the relevance of morphological structure. A basic observation is that contraction affects obstruents at the end of a stem, but not at the end of a word (Martin 1952; McCawley 1968; Vance 1987; Ito and Mester 1996). Stem-final /t/ in /bet-seki/ 'different seat' is contracted as in /bes-seki/, and the non-contracted form */betu-seki/ does not occur. On the other hand, word-final /t/ in {toku-bet}-seki 'special seat' is not contracted and /u/ is inserted as in {toku-betu}-seki ('{ }' indicate word boundaries). Contraction does not occur before the word boundary, e.g., *{toku-bes}-seki. See Ito and Mester (1996: 35–39) for more details.

However, again we find occasional inconsistencies of contractions, e.g., {san-kaku}-kei~{san-kak}-kei 'triangle'. At first glance it seems that contraction occurs before the word boundary in these cases, but Vance (1987: 162) points out that the contracted form is 'undoubtedly a secondary development fostered by vowel devoicing.' This implies that what occurs before the word boundary is devoicing and deletion of vowels, which results in contraction.

If devoicing accounts for the inconsistency of contraction at the word boundary, then, it might account for the one at the stem boundary, like /geki-ka/~/gek-ka/ discussed above as the second type of exceptions. Then we might want to know why the first type of exceptions, like /eki-ka/, not */ek-ka/, rejects devoicing/contraction. That devoicing accounts for the inconsistency of contraction in two-morpheme compounds needs further investigation "based on oral, spontaneous, and authentic data," as Labrune (2012: 34) concludes her section on vowel insertions and deletions.

5 Geminates in mimetics

Alongside the native and SJ vocabulary items so far discussed, there is a substantial class of mimetic, sound-symbolic items. This section provides an overview of the distribution and the function of geminates in this type of vocabulary, focusing on mimetic adverbs, which are by far more numerous than mimetic nominal adjectives (Hamano 1998: 25).

According to Hamano (1998: 25–28), mimetic adverbs consist of a root and a suffix, followed by the quotative particle /to/. The root has a canonical form of either CV or CVCV. The suffixes added to CV roots are either a coda nasal or a coda obstruent (a geminate). Thus a CV root such as /pu/ results in /pun-to/ 'a whiff', and /put-to/ 'spitting out'. Reduplication of the monosyllabic stems produces forms such as /pun-pun-to/ 'giving out a strong smell'. The same can be said of CVCV roots. The CVCV root, such as /pata/, turns up in forms with suffixes like /patan-to/ 'with a bang', /patat-to/ 'plop', /patari-to/ 'suddenly' and with reduplication like /patan-patan-to/ 'whap-whap'. Note that the suffix /ri/ attaches only to CVCV roots.[10]

5.1 Geminates as mimetic infixes

Focusing on mimetics with disyllabic roots, Nasu (2007: 47) classifies the occurrences of coda obstruents into two types: the infixation type and the suffixation type. In the former, geminates appear word-medially in the intensified adverbs (e.g., /pattari/ 'abruptly') and in the reduplicated forms (e.g., /passapasa/ 'dry, brittle'). In the latter they appear word-finally before the particle /to/ (e.g., /pa.tat-to/ 'suddenly').

Intensified adverbs take the form /CVC.CV.ri/, where the second C, in the coda position, is an infix (Kuroda 1965: 205–206). These adverbs are often, though not necessarily (cf. (11c)), related to reduplicated mimetic adverbs of the form /CVCV +CVCV/, as in (11), which gives rise to the analysis of /CVCV/ as a mimetic root of the intensified adverbs. So the form /battari/ in (11a) has the root /bata/ with the suffix /ri/ and the infix C (e.g., /baCta-ri/).

(11) Intensified adverbs (based on Vance 1987: 45)

a.	battari 'suddenly'	cf. batabata
	hakkiri 'clearly'	cf. hakihaki
b.	zamburi 'with a splash'	cf. zabuzabu
	bonyari 'vaguely'	cf. boyaboya
c.	sikkari 'firmly'	*sikasika
	yukkuri 'slowly'	*yukuyuku

10 Waida (1984) and Hamano (1998) list vocalic elements as suffixes alongside these three. Here we limit ourselves to these three suffixes, following Nasu (2007).

The infix consonant, underlyingly unspecified except as a consonant, is nasalized and surfaces as a coda nasal before a voiced consonant and before a sonorant as in (11b). On the other hand, an infix before a voiceless obstruent (11a) is totally assimilated to the adjacent consonant and surfaces as a geminate.

These are the same processes that we saw above in verbal inflection, that is, the case where a consonant followed by another consonant underlyingly (/CC/) surfaces either as a geminate or as a coda nasal, which are complementary in nature. However, while in verbal inflection the underlying /CC/ sequence occurs as a side effect of placing two morphemes in sequence, the infixed consonant in mimetics is added in order to create a geminate which expresses intensification.

Infixation, and therefore gemination, applies to the reduplicated forms as well, yielding /pik.ka.pi.ka/ 'very brightly' and /pi.kap.pi.ka/ 'very brightly' from the reduplicated form of /pika-pika/ 'brightly', for example. Based on a statistical study of the grammaticality judgment by 91 native speakers of Japanese, Nasu (2008: 72) points out that the HL-LL pattern (/pik.ka.pi.ka/) is most favored (favored by 83% of the participants) as an output emphatic form, followed by the LH-LL pattern (/pi.kap.pi.ka/) and least favored is the LL-HL pattern (*/pi.ka.pik.ka/), where H stands for a heavy syllable with an infix coda obstruent and L stands for a light syllable with no infix. From this, he concludes that the occurrence of the word-medial geminates obeys the condition on the balance of syllable weight between the two members of the reduplicated base and that the prosodic structure of a word-initial HL sequence is significantly preferred to an LH sequence.

The tendency in native Japanese phonology to favor an HL sequence to the one of LH in word-final position has often been pointed out in the literature (Tateishi 1989; Ito, Kitagawa, and Mester 1996; Kubozono 2003; among others). This tendency is observed in various phenomena in Japanese, including zuuzya-go formation (e.g., /gon.ta/ from /tan.go/ 'tango'), loanword truncation (e.g., /pan.ɸu/ from /pan.ɸu.ret.to/ 'pamphlet') and baby words (e.g., /kuk.ku/from /ku.tu/ 'shoes'); see Kubozono (this volume) and Ito and Mester (Ch. 9, this volume) for more details. The infixation in reduplicated mimetic forms shows the same tendency to prefer HL sequences over LH ones.

5.2 Geminates as mimetic suffixes

Geminates of the suffix type, which never occur in other lexical strata, are quite frequent in mimetics. They account for 90% of the 230 tokens according to the survey in Nasu (2007: 49). The list (12) exemplifies the forms with the same base occurring with different suffixes: a coda obstruent (12a), /ri/ (12b) and a coda nasal (12c). ('{ }' denotes a word boundary.)

(12) Mimetic suffixes with /pita/ and /doki/
a. {pita-t}-to ‘tightly’ {doki-t}-to ‘taken aback’
b. {pita-ri}-to ‘tightly’ {doki-ri}-to ‘startled’
c. {pita- n}-to ‘tightly’ {doki- n}-to ‘feeling a shock’

Following Waida (1984) and Hamano (1998), Nasu (2007: 51) argues that the three suffixes are obligatory elements for the mimetic vocabulary to surface. Of these three suffixes, coda obstruents (geminates) are the most productive and occur with most mimetic base forms, while the other two are more selective, occurring only with certain base forms. To take the base /boke/, for example, we find {boke-t}-to ‘absent-mindedly’ with a coda obstruent, but no *{boke- n}-to with a coda nasal, nor *{boke-ri}-to with /ri/. Nasu (2007: 48–51) argues that gemination is the default type of mimetic suffix, based on its wider range and greater frequency of occurrence in mimetics compared to the other two suffix types. He further argues that this preference for geminates is due to their semantic neutrality, asserting that each of the other two suffix types adds some symbolic meaning to the base form.

Observing that word-final gemination never occurs in the other lexical strata, one may naturally wonder if its appearance in the mimetic stratum has some good reason, perhaps a prosodic motivation. Nasu (2007: 55) claims this is due to the accent pattern specific to mimetics (see Nasu, this volume, for details about mimetic phonology). The prosodic pattern of the form in (12a) is represented as {pi (ta’t)}, where ( ) and { } denote foot and word boundaries, respectively, and an apostrophe indicates word accent, placed immediately after the accented vowel/mora. The prosodic pattern of the form in (12b), on the other hand, is represented as {pi (ta’ri)} and the one in (12c) as {pi (ta’n)}. In all these forms, the accent is on the final foot.[11] In order to place accent on the final foot (more strictly on the head of the foot) in the forms with no suffix, gemination is triggered. Take the base form /pita/, for example, for which four prosodic structures are possible: {pi.(ta’t)}, {pi’.(tat)}, {(pi’.ta)} and {(pi.ta’)}. The first structure, with a coda obstruent and accent on the final foot, is found to be the most well-formed because the other three candidates violate some crucial constraint pertaining to the prosodic structure of mimetics.[12] The peculiar characteristic of word-final appearance of geminates is thus required to maintain accent on the word-final foot, a prosodic structure seemingly specific to the mimetic vocabulary.

11 Similarly, the accent is on the final foot in reduplicated forms, e.g., [pi (tapi)(ta’t)] /pitapitat/, [pi(tapi)(ta’n)] /pitapitan/, [(pita)(ta’t)] /pitatat/, [(pita)(ta’n)] /pitatan/.

12 Of the three possible prosodic structures, [pi’.(tat.)], [(pi’.ta.)] and [(pi.ta’.)], the first two violate Align-Right{accented syllable, prosodic word}, which demands that every accented syllable stands in final position of the prosodic word. The third form [(pi.ta’.)] violates RhythmType=Trochee, which demands that feet have initial prominence (Nasu 2007: 53), violated here because the foot has accent on the final syllable, and not on the initial one.

Note that this prosodic structure can be reanalyzed if considered in a wider context.[13] The geminate-final mimetic form in question as well as other mimetic forms with an accent on the final foot is not used on its own in connected speech. In other words, {pi.(ta’t)}, {pi (ta’ri)} and {pi (ta’n)} are always accompanied by the particle /to/ when they are actually used in speech. Assuming that the particle is part of the prosodic word, the accented foot in these words is no longer in final position: {pi.(ta’t)-to}, {pi(’ta’ri)-to}, {pi(ta’n)-to}. This prosodic structure involves a sequence of heavy-light syllables word-finally, which is very commonly found in other lexical strata (see section 6 below and Kubozono, this volume). Seen in this light, what appears to be word-final gemination in mimetics can be understood in the same way as gemination in mimetic infixes mentioned in section 5.1 above, as well as gemination in loanwords to be discussed in section 6 below: namely, it is motivated to improve the prosodic structure of the output form.

In addition to this, we find some fundamental similarities to other lexical strata. In mimetics, voiced geminates are marked and prohibited (Nasu 2008: 70), just as in the native and SJ vocabulary. The word-final coda obstruent in the mimetic vocabulary seems not able to be linked to a following consonant (see section 2), but in practice we find it is licensed by linking to the onset of the following particle /to/.

In summary, geminates in the mimetic stratum do not arise in order to remedy an illegal /CC/ concatenation, as in the native stratum; rather, they occur as an emphatic infix word-medially, to produce a preferred prosodic sequence in reduplicated forms, and as a default suffix in word-final position to attain a certain prosodic structure. Gemination in the mimetic vocabulary is thus motivated by both semantic and prosodic factors.

6 Geminates in loanwords

6.1 Introduction

We have seen the distribution of geminate obstruents in three strata of Japanese – the native, SJ and mimetics vocabularies – and confirmed that gemination takes place for various purposes, to remedy phonotactic structure, for intensification, to show the integrity of a compound word, or to attain a certain prosodic structure. We now turn to loanwords, the lexical stratum in which geminate obstruents are extremely common (Kubozono, Ito, and Mester 2008: 959). In this section, we will consider why geminates are so common in loanwords and, more fundamentally, why they appear in this type of word at all. The question is whether they appear, as in other types of words, to remedy the phonotactic structure, for intensification, to

13 I owe the idea in this paragraph to Haruo Kubozono (personal communication).

show the integrity of the word, or to achieve a certain prosodic structure. One answer that seems widely accepted in the literature is that the source words, most of which are English words, contain something that makes Japanese speakers hear geminates. Even though English has no geminate consonants in its phonological system, something in the source string might trigger the perception of a geminate to Japanese *ears* during the adaptation of the English word into Japanese.

A number of perception studies have been conducted, using English words or English-like non-words as stimuli, with native speakers of Japanese as listeners (Hirozane 1992; Takagi and Mann 1994; Arai and Kawagoe 1998; Kubozono, Takeyasu, and Giriko 2013; and others). As reviewed in Kawahara (this volume), ratios of relative durations like C/V (the ratio between the target consonant and the preceding vowel) or C/C+V (the ratio between the target consonant and the preceding mora) are considered to be at work as primary cues in geminate perception, with several other factors as possible secondary cues. Each experimental participant listens to several tokens of a single form, and some listeners identify a geminate while others do not. Note here that the identification rate, i.e., total geminate responses over all the tokens, is in the range of 50% to 90% (Hirozane 1992: 18) when the target consonant is an unvoiced plosive in word-final position and the preceding vowel is short.

On the other hand, the occurrence of geminate obstruents in lexicalized loanwords is much more stable and consistent. When the target consonant is a voiceless plosive in the same environment as noted above, gemination is always observed in established loanwords, with no variability. This contrast in consistency between the variable perception of geminates in newly encountered strings and their invariable occurrence in lexicalized loanwords suggests that gemination cannot be attributed to perception alone.

Kubozono (2006: 1168), discussing the accentuation of loanwords in Japanese, argues that perception of the source forms influences the adaptation of accent, but even more importantly that “this perceptual process is already constrained by the prosodic system of the recipient language.” Native speakers of Japanese might think that they *hear* geminate obstruents in the source forms, but this perception process itself is influenced and constrained by the native phonology. To answer more completely the questions we asked above – why geminate obstruents appear in loanwords, and why they are extremely common in loanwords – we need to look at the distribution of geminate obstruents in loanwords and see how it can be accounted for in phonological terms.

In this section, we will first look at the distribution of geminate obstruents in loanwords in terms of segmental and contextual conditions. We will then give an overview of phonological research on this topic and consider two particular accounts in some detail. Finally, a critique and consideration of remaining problems will be offered.

6.2 Distribution of geminates in loanwords

6.2.1 Two basic conditions

Searching Nihon Hōsō Kyōkai (1987) for borrowings from English, Kitahara (1997: 217) reports 2,476 words, among which 420 contain the small-sized letter 'tsu' representing geminate obstruents.[14] His analysis of these 420 words leads to the following two observations:

(13) Two basic conditions on gemination in loanwords (Kitahara 1997: 218)
  a. Segmental condition: The consonants that geminate (target C) are obstruents and not sonorants.
  b. Contextual condition: The vowel preceding a target consonant is short, e.g., /a/, /e/, /i/, /o/, /u/; no diphthongs, no long vowels, and no epenthetic vowels.[15]

The segmental condition in (13a) can be exemplified by geminated plosives (14a), fricatives (14b), and affricates (14c), whereas forms with sonorants in the relevant position (14d) have no geminates. The vowels and consonants of foreign source words are transformed to conform to Japanese phonology, so English tense vowels, for example, are changed to corresponding long vowels in Japanese (see Ito and Mester Ch. 7, this volume, for more details).

(14)	a.	kyap.pu	'cap'	pet.to	'pet'	pik.ku	'pick'
	b.	kyas.syu	'cash'	pah.hu	'puff'	guz.zu	'goods'
	c.	kyat.ti [tʃi] 'catch'	kyat.tu [tsɯ]	'cats'	zyaz.zi	[dʒi]	'judge'
	d.	ha.mu	'ham'	pen	'pen'	pi.ru	'pill'

Examples in (15) illustrate the second observation, the contextual condition, showing the complete absence of gemination of the same plosives following long vowels (15a) and diphthongs (15b).

(15)	a.	poo.pu	*poop.pu	'pope'
		hii.to	*hiit.to	'heat'
		ruu.ku	*ruuk.ku	'Luke'
	b.	sai.to	*sait.to	'site'
		rei.ku	*reik.ku	'lake'
		paa.ku	*paak.ku	'park'

14 We restrict our attention to geminate obstruents and do not discuss nasal geminates, such as /tonneru/ 'tunnel'.

15 Epenthetic vowels are vowels with no corresponding segments in the source.

While the conditions in (13) accurately predict the occurrence of geminates, they also wrongly predict the occurrence of geminates in words where there are actually none. In the same study where 420 loanwords with geminates were identified, Kitahara (1997: 218) counts over 300 loanwords where obstruents do not geminate even though they meet the conditions in (13). In order to avoid overgeneration of geminate consonants, additional segmental and contextual conditions are needed to further constrain the gemination.

6.2.2 More on segmental conditions

Since the native phonology of Japanese does not allow gemination of voiced obstruents (Kuroda 1965), it is natural to assume that loanwords will show the same limitations. However, we do find voiced geminates in loanwords. All the words in (16) come from sources with a target C in word-final position preceded by a short vowel. When the target C is a voiceless obstruent, it is almost always geminated as in (16a), while gemination of voiced obstruents is permitted in some words as in (16b), but is avoided in others as in (16c).

(16)	a.	kap.pu	'cup'	kat.to	'cut'	bak.ku 'back'
	b.	su.nob.bu	'snob'	bed.do	'bed'	bag.gu 'bag'
	c.	pa.bu	'pub'	a.do	'ad'	ba.gu 'bug'

Ito and Mester (1999) account for the difference of gemination between the forms in (16b) and in (16c) as a difference of lexical strata (Foreign vs. Alien). The assimilated foreign items do not allow voiced geminates as in /pabu/ in (16c), while the unassimilated alien items allow the voiced geminates, as in /beddo/ in (16b). Voiced geminates are marked, so avoided by degemination in the assimilated foreign vocabulary.

Some of the forms with voiced geminates are observed to undergo devoicing, showing free variation between voiced and voiceless geminates: /beddo/~/betto/ and /baggu/~/bakku/. Nishimura (2003) claims and Kawahara (2011) experimentally confirms that devoicing primarily occurs when the form contains more than one voiced obstruents within one morpheme, which is due to an OCP effect. See Vance (this volume) for this constraint, which is also known as Lyman's Law. Compare /beddo/ 'bed' with /heddo/ 'head'. The former with an initial voiced obstruent permits a devoiced variant, i.e., /betto/, while the latter with an initial unvoiced obstruent shows no variation. In the former, the marked pattern of voiced geminates is avoided by devoicing, but this process occurs only when an OCP is violated. See Kawahara (2012) for a summary of theoretical analyses of this issue.

Hirayama (2008: 79–81) reports asymmetry in gemination of voiced plosives, based on gemination rates of various places of articulation in word-final position

(see also Shirai 1999).[16] According to this report, the labial [b] is the most resistant to gemination at 11–23%, the coronal [d] is the least, with 58–83%, and the dorsal [g] is in between, 42–55%.

Among other categories of obstruents, /z/ is likely to be realized as an affricate when geminated, as seen in [bad.ʤi] 'badge' (/bazzi/) and [gud.ʣɯ] 'goods' (/guzzu/). Maruta (1999) counts 21 instances with [ʤ] and [ʣ], and all are geminated.[17] The behavior of voiceless fricatives is another puzzle for the generalizations in (13). Two voiceless fricatives, /s/ and /h/ (realized as [ɸ]), are not usually geminated, but /sj/ (*sy* [ʃ]) is quite freely geminated. Some examples are given in (17). It is a mystery why these phonetically similar segments show such differences in gemination. This will be discussed again in section 6.3.5.

(17)	a.	/s/	bo.su	'boss'	ku.ra.su	'class'
	b.	/h/	ta.hu	'tough'	gu.ra. hu	'graph'
	c.	/sj/	bus.syu	'bush'	kyas.syu	'cash'

6.2.3 More on contextual conditions

So far we have considered those forms with a target C in word-final position in the source word: e.g., for /kjap.pu/ in (14a), the source word is 'cap,' with a target C in word-final position. When we look at a target C in word-medial position, however, we find two contextual conditions regulating the occurrence of geminates.

Let us first compare /bat.to/ 'bat' and /ba.to.raa /'butler'. Both forms have voiceless plosive /t/ as a target C, but only /bat.to/ has a geminate. Additional examples showing this contrast are given in (18).

(18)	kyap.pu	'cap'	cf. kya.pu.ten 'captain'
	hit.to	'hit'	cf. hi.to.raa 'Hitler'
	dok.ku	'dock'	cf. do.ku.taa 'doctor'

The forms in (18) may give the impression that gemination occurs only in the target C in word-final position, but consider the geminated form /bat.taa/ 'batter', where the target /t/ appears word-medially. The difference between /ba.to.raa/ 'butler' and /bat.taa/ 'batter' may be due to different contextual conditions on the target C in the source word. In 'butler', /t/ is followed by another consonant in the source, while /t/ in 'batter' is in an intervocalic position, immediately followed by a vowel. The latter case is more complicated since gemination seems optional and

16 The wide range of percentage is due to the fact that three different surveys were used.
17 Maruta (1999: 104) includes three instances with [ʣ] (e.g., [gud.ʣɯ]) as fricatives, but we count them as instances of affricates.

variable. Compare, for example, /bat.taa/ 'batter' with /ba.taa/ 'butter', where the target /t/ is in inter-vocalic position in both sources, but one becomes geminated and the other does not. (19) gives additional examples of such contextually similar pairs, those with gemination (19a) and those without (19b).

(19)	a.	hap.pii	'happy'	mot.too	'motto'	sak.kaa	'soccer'
	b.	pa.pii	'puppy'	re.taa	'letter'	ti.kin [tʃikin]	'chicken'

In short, we find two types of contextual conditions on gemination in the source medial position: When the target C is followed by another consonant in the medial position in the source, gemination does not take place as in (18). When the target C is followed by a vowel in the source, gemination is optional as in (19).

Another context where the conditions in (13) make wrong predictions is where there is a voiceless consonant cluster in word-final position in the source word. When the source word ends in the consonant cluster /-ks/, /k/ is always geminated as in (20a), but when the source word ends in /-kt/, /-pt/ or /-sk/, gemination does not occur as in (20b). Gemination in these forms will be discussed in section 6.3.6.

(20)	a.	/-ks/	tak.ku.su 'tax'	mik.ku.su 'mix'
	b.	/-kt, -pt, -sk/	ta.ku.to 'tact'	kon.se.pu.to 'concept' ta.su.ku 'task'

At first glance, gemination looks like a simple phenomenon which applies to an obstruent when it is preceded by a short vowel, as stated in (13), but a closer examination reveals a more complicated situation. In this section, we have seen three contextual conditions in the source that produce variability in gemination: word-medial CC vs. VCV environments, and word-final clusters of voiceless obstruents.

6.3 Phonological accounts of gemination in loanwords

6.3.1 Why does gemination occur?

In order to account for the occurrence and non-occurrence of geminates in loanwords, two types of analyses have appeared across several theoretical frameworks: one makes the input source primary, and another focuses on the demands of the output. See Kang (2011) for detailed discussion on the perception of the foreign input.

The input-oriented analysis claims that gemination occurs because there is something in the source forms that motivates gemination. As Katayama (1998: 82) sums up early views, "one general insight shared by several researchers is that word-final gemination is the result of an attempt to keep the original closed syllable (Ohye 1967; Ohso 1971; Kunihiro 1963; Lovins 1973)." More recently, in the OT frame-

work, Tsuchida (1995) and Katayama (1998) agree that word-final gemination occurs to retain the closed syllable structure of the source words. It is also claimed that the ambisyllabicity of the intervocalic consonant (/p/ in 'happy') creates a closed syllable in the word-medial position and motivates gemination (Lovins 1973; Katayama 1998).

On the other hand, the output-oriented analyses, like Ono (1991) and Kawagoe (1995) in a rule-based approach, and Kitahara (1997) and Kubozono, Ito, and Mester (2008) in the OT framework, claim that the prosodic structure of the output forms primarily motivates gemination: Although the source forms are input to the adaptation process, they are heavily constrained by Japanese phonology and undergo a nativization process.

The idea of output-oriented approach finds phonetic support in Best et al. (1988), who claim that speech perception is highly constrained by the phonological structure of the listener's native language. Investigation on loanword accentuation in Kubozono (2006) also concludes that the pitch-shape of loanwords is constrained by the prosodic system of the recipient language. Thus below in this chapter we will presume that source forms are constrained by the recipient phonology and not directly accessible to the borrowing process. We now turn to a detailed consideration of two analyses in the output-oriented approach, using them to re-examine various cases that were introduced above.

6.3.2 Kitahara (1997)

6.3.2.1 Motivation for gemination

Whether the motivation to geminate comes from source or output forms, it is important to clarify what the input forms to loanword adaptation are. Kitahara (1997: 214) notes that for most previous analyses, input forms are assumed to be the pronunciations of the source language, with or without perceptual adjustments, but their exact nature is not clear. As Katayama (1998: 1) criticizes, "given the lack of independent evidence, it has been stipulative to select one form as the input over other possibilities."

Kitahara (1997: 214) addresses the input problem by calling upon the model of loanword phonology proposed in Silverman (1992: 293), which posits a Perceptual and an Operative level and specifies that the perceived acoustic signal is constrained by the segmental and tonal inventories of the borrowing language (see Kubozono, this volume, for a general discussion of loanword phonology). Kitahara posits the lexicon between the two levels in Silverman's model. Input forms, pronunciations of the source language, are forms at the Perceptual level. The outcome of the Perceptual level is stored in the lexicon, where lexical forms contain only unpredictable information. Predictable information is supplied by the phonology, which works at the Operative level. Thus, for Kitahara, input forms are the lexical representations of unpredictable information.

Phonology operates on the lexical forms, providing them with predictable information. Epenthetic vowels are predictable (Ura 1995), and most occurrences of geminates are assumed to be predictable. Word accent is usually on the head of the syllable containing the antepenultimate mora, or on the head of the syllable containing the penultimate mora in disyllabic words. This is the default accentuation, which is thus predictable (McCawley 1968). In Kitahara's model, phonology gives a systematic account of these predictable properties, while pushing the forms with unpredictable properties outside the phonological operation. Building on these models he claims that gemination and default accent are related, and that gemination occurs to attain default accentuation (Kitahara 1997: 222). Therefore, gemination in forms with accent on non-default positions (/sa'.mit.to/ 'summit', /e's.sen.su/ 'essence', /do'.ku.taa/ 'doctor') are not the concern of the phonology.

The OT tableau in (21) illustrates that the best interaction between gemination and prosodic structure (foot and syllable) that achieves default accentuation is the one that comes out as a winner; only candidates with default accentuation are considered here. Two constraints are relevant here: Align-R demands that the right edges of each foot and syllable be aligned,[18] and Fill-μ forbids any mora not in the input (such as epenthetic vowels and geminates) to be added to the output form. In (21) and the following tableaux, the following notations are used: L and H denote light and heavy syllable, respectively; "h)h" means that the second mora of a heavy syllable is not parsed into foot structure; "*" and "**" respectively show one and two violations of the relevant constraint; "!" means that the violation is crucial; ( ) indicates foot boundaries; an apostrophe indicates the position of accent; and ⇒ denotes the selected form.

(21) Word-medial gemination

	/sakaa/ 'soccer'	Align-R	Fill-μ
a.	⇒ sa'k.kaa (H')(H)		*
b.	sa'.kaa (L'h)h	!*	

The geminated form in (a) is selected as the optimal output because rival candidate (b) violates Align-R. While the right edge of syllable /sak/ aligns with the right edge of the foot /sak/ in (a), the right edge of the syllable /sa/ does not align with the right-edge of any foot in (b), e.g., /(sa.ka)a/. Since Align-R, which ensures proper prosodic structures representing default accentuation, is ranked higher (more important) than Fill-μ, the geminated form surfaces. The occurrence of geminates in this system is the result of a quest for the default accentuation.

18 Following Suzuki (1995), Kitahara assumes a bisyllabic or bimoraic foot.

6.3.2.2 Some problems

There are two problems with this analysis. One, also noted by Kitahara (1997: 229), is that there is a second possible default accent pattern for Japanese loanwords, namely, accent on the pre-antepenultimate mora, as in /sa'.mit.to/ 'summit', /e's.sen.su/ 'essence' and /do'.ku.taa/ 'doctor,' mentioned above. If this accentuation is also a default one, gemination and no gemination in these forms ought to be predicted and accounted for in the phonology, which this model does not do.

The other problem comes from the case of word-final gemination, like /kat.to/ 'cut', as illustrated in (22). The relevant constraint here is Align-stem, which demands the alignment of the right edge of the stem to the right edge of a syllable.

(22) Word-final gemination Kitahara (1997: 221)

	/kat/ 'cut'	Align-stem	Fill-μ
a.	⇒ ka't.to (H')L		**
b.	ka'.to (L'L)	!*	*

The candidate in (a) wins because it satisfies Align-stem, while the candidate in (b) does not. Noteworthy here is the reference to the stem-edge of the input structure: Candidate (b) is ruled out not because of its word-final prosodic structure (LL), but because it does not preserve the structure of the input stem. In word-medial position, as in (21), well-formed output prosodic structure is the key to the selection of surface form. In word-final position, in contrast, the output prosodic structure is not at issue, but the alignment to the input stem is. Thus, the gemination process is motivated by different constraints in word-medial and word-final positions.

Kitahara (1997) claims that gemination occurs to preserve the default accentuation which is represented in the output prosodic structure, but (22) shows that word-final gemination has nothing to do with the output prosodic structure and, as such, word-final gemination has nothing to do with default accentuation. In order to pursue the claim of gemination to preserve default accentuation, gemination in word-final position needs further explanation. This point makes clear the difference between the analyses of Kitahara (1997) and Kubozono, Ito, and Mester (2008). As will be seen below, it is the prosodic structure of LL at word-final position that is disfavored in the Japanese phonology according to Kubozono, Ito, and Mester (2008).

6.3.3 Kubozono, Ito, and Mester (2008)

6.3.3.1 Motivation for gemination

Kubozono, Ito, and Mester (2008: 959) propose an analysis with three claims: "First ... geminate consonants are universally more marked than singletons. Second, despite

this markedness, obstruents can be geminated in order to improve prosodic well-formedness. Third, gemination does not occur when it is not motivated: that is, it is blocked either if it will not improve prosodic well-formedness or if it would produce a structure that is banned in native phonology."

The first claim is embodied as a constraint *Gem (no geminates), which is similar to constraints in other OT-based approaches, like constraint Fill-μ in Kitahara (1997), discussed above. Both *Gem and Fill-μ disfavor geminate consonants, but Fill-μ disallows geminates because they are not in the input, while *Gem disfavors geminates as universally marked structures. In Kubozono, Ito, and Mester (2008), whether the geminate is in the input or not is of no concern. Universal markedness is the reason why all geminates are disfavored.

The second claim explains why gemination occurs in a certain context, while the third explains why gemination does not occur in a certain context. In either case, it is the creation of well-formed prosodic structure that motivates either gemination or no gemination. Prosodic Form (ProsForm) monitors the favored prosodic structures assumed for Japanese word-final position, which are Heavy-Light (HL) and Heavy-Heavy (HH) structures; the disfavored sequences are Light-Heavy (LH) and Light-Light (LL).[19] ProsForm is the constraint that primarily triggers consonant gemination in loanwords.

Some other relevant constraints for gemination are *VoiGem (no voiced geminates), *σμμμ (no superheavy, i.e., trimoraic, syllables), and *μμ'μ]PrWd (no penultimate accent); see Kubozono (this volume) and Ito and Mester (Ch. 9, this volume) for details about the second constraint. The last one applies to trimoraic or longer words, prohibiting accentuation such as */ba.na'.na/ 'banana'. All these constraints are motivated in native phonology, and are not specially stipulated to account for loanword phonology. In the analyses below, constraints are ranked as follows:

(23) *μμ'μ]PrWd ≫ *VoiGem, *σμμμ ≫ ProsForm ≫ *Gem

6.3.3.2 Analysis of word-final gemination

Comparing the three tableaux below, (24a) shows a form with a short vowel followed by a voiceless obstruent, (24b) shows a short vowel and a voiced obstruent and (24c) shows a long vowel followed by a voiceless obstruent, all with the target consonants occurring at the end of the source word.

19 The claim that HH and HL sequences are favored in word-final position, while LH sequence is disfavored there, is based on prosodic tendencies observed in various phenomena in Japanese, two of which are the *zuuzya-go* formation (Ito, Kitagawa, and Mester 1996) and the process of loanword truncation (Ito 1990; Kubozono 2003). Light-light (LL) syllable sequence is not mentioned in the definition of constraint Prosodic Form in Kubozono, Ito, and Mester (2008: 958), but in their Note 7 (p. 971) they mention that LL sequence violates ProsForm, so we include it here.

(24) Word-final gemination and non-gemination

a.

rick	*VoiGem	ProsForm	*Gem
⇒ .rik.ku.			*
ri.ku.		*!	

b.

rig	*VoiGem	ProsForm	*Gem
rig.gu.	*!		*
⇒ .ri.gu		*	

c.

leak	*σμμμ	ProsForm	*Gem
.riik.ku.	*!	*	*
⇒ .rii.ku.			

In (24a) a geminated form surfaces because the rival form without a geminate has LL prosodic structure, which is banned by ProsForm. In (24b), on the other hand, the ungeminated form surfaces in spite of its LL prosodic structure because the geminated rival form violates higher-ranked *VoiGem. In (24c), the geminated form is beaten because it has a trimoraic syllable, which violates *σμμμ, the highest-ranked constraint. Remember that (13b) states the vowel preceding the target C must be short, never a diphthong nor a long vowel; this condition is instantiated as a constraint against a superheavy syllable, *σμμμ.

In (24b), the ungeminated form surfaces but, as mentioned in section 6.2.2, gemination of voiced obstruents is often observed in loanwords. As seen in (25), reranking the two constraints *VoiGem and ProsForm predicts that the geminated form surfaces.

(25)

dog	ProsForm	*VoiGem	*Gem
⇒ .dog.gu.		*!	*
.do.gu.	*		

6.3.3.3 Analysis of word-medial gemination

Forms like /sak.kaa/ 'soccer' in (21) have a geminated target C which occurs in word-medial position in the source word. As illustrated in (26), the rival form */sa.kaa/ without a geminate has a disfavored word-final structure of LH, so the geminated form with HH structure surfaces.

(26) Word-medial gemination

soccer	ProsForm	*Gem
⇒ .sak.kaa.		*
.sa.kaa.	*!	

The absence of word-medial gemination in words like /do.ku.taa/ 'doctor' (not */dok.ku.taa/) in (18) is a problem for many analyses. Since the target consonant /k/ is in a coda position and preceded by a short vowel, gemination is expected. Kubozono, Ito, and Mester (2008) claim that what is crucial here is the prosodic structure in word-final position. Both candidates /do.ku.taa/ and */dok.ku.taa/ have the same word-final sequence LH, which is prosodically undesirable. Since gemination does not improve the word-final structure, and since geminates are marked, the ungeminated form /do.ku.taa/ is selected as an optimal output form.

(27) Word-medial non-gemination (Kubozono, Ito, and Mester 2008: 964)

doctor	ProsForm	*Gem
.dok.ku.taa.	*	*!
⇒ .do.ku.taa.	*	

Two types of word-medial gemination and no gemination have been discussed in this section. The one with gemination has only one consonant word-medially in the input, like /sakaa/, and the other, with non-gemination, has two consonants word medially in the input, like /doktaa/. In the input-oriented approach, both gemination and non-gemination at word-medial position are often considered in relation to syllable structure. Gemination occurs when input forms have coda consonants. This approach, typically represented by Katayama (1998: 136), is problematic, because whether the medial consonants are in coda position or not is decided by the occurrence of geminates in the surface forms. If no medial geminate appears, then no coda consonant is assumed in the input forms. Thus, in this approach, the argument becomes circular.

This problem does not concern the output-oriented approach, as proposed in Kubozono, Ito, and Mester (2008), since whether gemination occurs word-medially or not depends on output prosodic structure. Gemination occurs to preserve the prosodic well-formedness in word-final position; if all candidates are equally bad, the default winner will be the one without gemination. One problem with this proposal comes from words like /pa.pii/ 'puppy' with one medial consonant but no gemination. This will be discussed in section 6.4.

6.3.4 The effect of onset clusters on gemination

As mentioned in section 3.2, voiced obstruents and fricative /h/ ([ɸ]) usually do not geminate, but when the input form has a complex onset, they become less resistant to gemination, as seen in (28) (first reported by Ohye 1967: 116).

(28)	hu.rog.gu ‘frog’	vs.	ro.gu ‘log’
	do.rag.gu ‘drug’	vs.	ra.gu ‘lag’
	su.nob.bu ‘snob’	vs.	no.bu ‘knob’
	su.tah.hu ‘staff’	vs.	ta.hu ‘tough’

Kubozono, Ito, and Mester (2008: 967) claim that the answer to this phenomenon lies in the accent structure of /hu.ro’.gu/ and /do.ra’.gu/, where the accent on the penultimate mora violates the constraint *μμ’μ]PrWd. The ungeminated form produces these marked accent structures, so the geminated form surfaces, as in (29a). On the other hand, when the onset of the input form is simplex, the ungeminated form surfaces as in (29b).

(29) Forms with complex onsets showing gemination (Kubozono, Ito, and Mester 2008: 968)

a.

frog	*μμ’μ]PrWd	*VoiGem	ProsForm	*Gem
⇒ .hu.ro’g.gu.		*		*
.hu.ro’.gu.	*!		*	

b.

log	*μμ’μ]PrWd	*VoiGem	ProsForm	*Gem
.ro’g.gu.		*!		*
⇒ .ro’.gu.			*	

In forms with complex onsets, an interesting interaction between accent change and consonant gemination is observed (Kubozono, Ito, and Mester 2008: 968). The form /hu.ra’g.gu/ ‘flag’ shows two variant forms: /hu’.ra.gu/ and /hu.ra.gu^{0}/, where superscript /^{0}/ denotes that the word is unaccented. It is claimed that these forms surface to avoid the marked penultimate accent structure, as illustrated in (30). In order to deal with this variation, a new constraint is introduced: Faith-accent (Faith-acc) (=A vowel which is accented in the input must be accented in the output). It is called “faithfulness constraint” because it demands that forms be faithful to the input (Prince and Smolensky 1993). A constraint ranking of Faith-acc ≫ *VoiGem selects the geminated and accented form (30a), while the reverse ranking

of *VoiGem ≫ Faith-acc selects the ungeminated forms, with either deaccenting or accent shifting (30b).

(30) Forms with complex onsets showing various accentuations (Kubozono, Ito, and Mester 2008: 969)

a.

flag	*μμ'μ]PrWd	Faith-acc	*VoiGem	ProsForm	*Gem
⇒ .hu.ra'g.gu.			*		*
.hu.ra'.gu.	*!			*	
.hu.ra.gu^{0}.		*!	*		
.hu'.ra.gu.		*!		*	

b.

flag	*μμ'μ]PrWd	*VoiGem	Faith-acc	ProsForm	*Gem
.hu.ra'g.gu.		*!			*
.hu.ra'.gu.	*!			*	
⇒ .hu.ra.gu^{0}.			*	*	
⇒ .hu'.ra.gu.			*	*	

In (30a) the marked accent structure was avoided by gemination, while in (30b) it is avoided by accent change, either by shifting the accent to the antepenultimate mora or by deaccenting the word. Forms that are both geminated and deaccented, e.g., /hu.rag.gu^{0}/, never appear.[20] The geminated form surfaces when the accent structure of the input is preserved (30a), and the two ungeminated forms surface when the native constraint of *VoiGem is preserved (30b). The result of (30a) is more faithful to the input's accent, while the one of (30b) conforms more to native phonotactics. The reranking of the two constraints reflects different levels of nativization of loanword forms.

6.3.5 Word-final sonorants and fricatives

Obstruents undergo gemination but sonorants do not, as mentioned in (13a). Among the three sonorants illustrated in (14d), the alveolar nasal [n] in the source word-final position surfaces as a coda nasal unless the loans are from French, e.g., /kannu/ 'Cannes' (Peperkamp, Vendelin, and Nakamura 2008), while [m] and [r]

20 However, some geminated forms with accent on the pre-antepenultimate mora are observed, e.g., /su'.rip.pa/ 'slipper', /hu'.rip.pu/ 'flip', /to'.rik.ku /'trick'.

surface as onset singleton consonants (e.g., /ha.mu/ ‘ham’, /pi.ru/ ‘pill’). The non-gemination of sonorants is often discussed together with the non-gemination of word-final fricatives.[21] As seen in (17a) and (17b), fricatives /s/ and /h/ in word-final position are usually not geminated. Kitahara (1997: 226) claims that non-gemination of /s/ and /h/ as well as that of /r/ is due to their syllabicity: these segments make their own syllables and are thus not syllabified as coda consonants in Japanese. Since they are not in coda position, there is no motivation for gemination in his analyses.[22]

Kubozono, Ito, and Mester (2008: 958) claim that the absence of gemination of /s/ and /h/ is caused by the loss of syllabicity of the following epenthetic vowels, which is called extraprosodicity.[23] This claim is supported by the accent pattern of the syllables /su/, /hu/ and /ru/ in word-final position, which do not behave like other light syllables but like part of a heavy syllable. Thus, /ku.ra.su/ ‘class’ in (17a) is represented as /ku.ras.<u>/, /ta.hu/ ‘tough’ (17b) as /tah.<u>/ and /pi.ru/ ‘pill’ (14d) as /pir.<u>/, where < > indicates extraprosodicity. The final vowels are not counted prosodically and the segments /s/, /h/ and /r/ are syllabified as codas of the preceding syllables, making them prosodically heavy. The tableau in (31) shows that the ungeminated form is selected over the geminated one, because the latter has a superheavy syllable created by extraprosodicity.

(31) Forms with an extraprosodic element (Kubozono, Ito, and Mester 2008: 968)

tough	*μμ’μ]PrWd	*σμμμ	ProsForm	*Gem
.tah.h<u>		*!	*	*
⇒ .tah.<u>			*	

Fricatives /s/ and /h/ are not geminated, but another fricative /sj/, realized as [ʃ], is usually geminated as in (17c).[24] For both the syllabicity and extraprosodicity accounts, gemination of /sj/ ([ʃ]) is a mystery. In Kubozono, Ito, and Mester (2008:

21 Concerning the nasal [m], no discussion is found in Kitahara (1997) or Kubozono, Ito, and Mester (2008). Ono (1991: 82) treats [m] as an extraprosodic element, together with [s], [ɸ] and [r].

22 The syllabicity of [s] and [ɸ] is supported by their high sonority and by devoicing of the following vowel. Hashimoto (1993, cited in Katayama 1998: 88) claims the appearance of devoiced vowels on spectrograms is similar to that of fricatives. This similarity can allow devoiced vowels to merge with fricatives, which are then syllabic.

23 In Kubozono, Ito, and Mester (2008, 958), [rɯ] is assumed to lose its syllabicity like [sɯ] and [ɸɯ] in the accent behavior, but the absence of gemination of [r] as in /be.ru/ ‘bell’ (not */ber.ru/) is not discussed in this connection.

24 The gemination of [ʃ] depends on the following vowel. It usually geminates when followed by epenthetic [ɯ], like [bu.raʃ.ʃɯ] ’brush,’ but it does not geminate when followed by epenthetic vowel [i], like [bu.ra.ʃi] ‘brush’. The latter is probably a relic of an old pattern of borrowing that is no longer productive.

962), the different behavior of forms with final [ʃɯ] is assumed to be the result of lack of extraprosodicity, but evidence for this assumption has been lacking.

Recently, however, Matsui (2012: 67) presents acoustic and perceptual evidence for the difference between word-final [ʃɯ] and [sɯ]. In the former, where the final vowel does not behave as extraprosodic and gemination is observed, a formant transition appears from [ʃ] to [ɯ], while in the latter, where extraprosodicity is assumed and no gemination occurs, no formant transition is observed. From this Matsui concludes that a formant transition serves as a perceptual cue to gemination by marking the end of frication.

Gemination of the fricative /h/ ([h]) is prohibited in the native phonology, but occurs in loanwords like /bah.ha/ 'Bach'. Kubozono, Ito, and Mester (2008: 962) interpret this as showing "that loanwords are sensitive to phonological (markedness) constraints to a lesser extent than native words." This is another instance that requires phonetic and phonological explanations.

6.3.6 Word-final consonant clusters

Kubozono, Ito, and Mester (2008) claim that whether gemination occurs or not in the word-final consonant clusters shown in (20) results from a difference in the extraprosodicity of the final vowel. Compare the geminated form /tak.ku.su/ 'tax' in (20a) with the ungeminated form /ta.ku.to/ 'tact' in (20b). The former word ends in [sɯ], the latter in [to]. With the notion of extraprosodicity, the former will be syllabified as /tak.kus.<u>/ with a final extraprosodic vowel. This is a favored prosodic structure involving HH in final position. With no gemination, in contrast, it would be syllabified as /ta.kus.<u>/, which is a disfavored prosodic structure involving LH. Thus, the geminated form has a better prosodic structure and is thus selected through constraint interaction. This account does not apply to /ta.ku.to/ 'tact' in (20b), however. This word has no extraprosodic element, and is thus syllabified as /ta.ku.to/ with a word-final LL structure. While this is not a preferred prosodic structure, geminating /k/ would still result in a word-final LL structure, i.e., /tak.ku.to/. Since gemination does not improve the prosodic structure, the ungeminated form is selected in this case.

What matters for Kubozono, Ito, and Mester (2008) is the prosodic structure in word-final position, while it is the prosodic structure of the whole word that is relevant for Kitahara (1997). In the latter account, the geminated form /tak.ku.su/ 'tax' is analyzed as /(ta'k.)(ku.su)/ (H)(LL), and the ungeminated form as /(ta'.ku.)su/ (LL)L. The latter form is rejected because word-final /su/ is stipulated as an underlying vowel to represent the syllabicity of /s/ and is not properly incorporated into a foot structure. On the other hand, the ungeminated form of /ta.ku.to/ 'tact', i.e., /(ta'.ku.)to/, also has an unfooted syllable, but this time it is well-formed because the final /o/ does not underlyingly exist, and thus need not be included in the foot structure.

The syllabicity of /s/, which is interpreted as /su/ underlyingly, plays a crucial role in this analysis.

In both Kubozono, Ito, and Mester's (2008) and Kitahara's (1997) analyses, the different gemination behavior shown by the final consonant clusters /ks/, /kt/, and /pt/ is assumed to result from the different phonetic qualities of word-final [s] and [t].

6.4 Summary of two analyses, and some residual problems

This section has discussed two analyses of the distribution of geminates, one by Kitahara (1997) and the other by Kubozono, Ito, and Mester (2008). What is common to both is the contention that it is not the source form per se but the structure of the Japanese phonology that directly motivates gemination in loanwords. They agree that gemination occurs to preserve a well-formed output prosodic structure, but they differ in what the well-formed prosodic structure is. In Kitahara (1997), a well-formed structure is the one that best represents default accentuation, while in Kubozono, Ito, and Mester (2008) it is the structure with a word-final HH or HL sequence.

The difference between these two analyses is seen in the account of word-medial non-gemination of /k/ in the form /do'.ku.taa/ 'doctor'. In Kitahara (1997), gemination versus non-gemination in forms with non-default accentuation is assumed to be unpredictable and to be outside of phonological operations. Since /do'.ku.taa/ is accented on a pre-antepenultimate mora, it represents a non-default accent pattern, and hence, its non-gemination is outside the scope of a phonological analysis.

In Kubozono, Ito, and Mester (2008), on the other hand, non-gemination of /k/ in /do'.ku.taa/ is attributed to the relative markedness of the geminated form. Since geminated (/dok.ku.taa/) and ungeminated forms (*/do.ku.taa/) have the same prosodic structure in word-final position, the ungeminated form will surface because gemination does not make the output less marked in their analysis. Gemination does not improve the well-formedness of the output structure and is, hence, not motivated. In sum, non-gemination of the form /do.ku.taa/ is assumed to be phonologically unaccounted for in Kitahara (1997), but is given phonological explanation in Kubozono, Ito, and Mester (2008). The difference between these two analyses is that gemination is closely related with the default accentuation in the former, while it is connected to the markedness of prosodic structure in the latter.

While these output-oriented accounts give us a new perspective on gemination in loanwords and offer new explanations for many traditionally unsolved problems, there are still several issues to be addressed in future research. One issue concerns the non-gemination of [r] (/be.ru/ 'bell'). Kitahara (1997) and Kubozono, Ito, and Mester (2008) propose the notions "syllabicity" and "extraprosodicity", respectively, to explain the non-gemination of [s] and [ɸ] in word-final position (section 6.3.5). These notions cannot account for the behavior of [r], however, since this sound

behaves differently from the two fricatives. While [s] and [ɸ] often geminate word-medially, [r] never geminates, either word-medially or word-finally.[25] Compare /be.rii/ (*/ber.rii/) 'berry' with /es.see/ 'essay' and /baɸ.ɸaa/ 'buffer'. Neither syllabicity nor extraprosodicity can thus account for the non-gemination of [r] in word-medial position.

Another problem concerns the occasional gemination of voiced obstruents. As presented in section 6.2.2, some forms readily geminate (e.g., /bag.gu/ 'bag'), but others resist the process (e.g., /ba.gu/ 'bug'). Gemination of /g/ violates *VoiGem, a constraint that is undominated in native phonology. Kubozono, Ito, and Mester (2008: 960) propose to account for the difference between /bag.gu/ and /ba.gu/ by reranking the two markedness constraints, *VoiGem and ProsForm. However, the reranking assumed here seems to contradict Ito and Mester's (1995: 183) claim that reranking should be limited to the faithfulness constraints as against markedness ones.[26] Further research will be needed to elaborate constraints and their relative rankings.

The last problem, not clearly discussed in either analysis, is word-medial non-gemination in forms like /pa.pii/ 'puppy'. In Kitahara (1997: 228), such forms are claimed to have lexically marked accent, but the reason why they do not geminate is not clear. In Kubozono, Ito, and Mester (2008), on the other hand, word-final LH structures violate ProsForm. Without demotion of this constraint, these forms should not surface. A reanalysis of the data in Kitahara (1997: 217, 219) reveals that there are 44 instances involving gemination (HH) and 39 instances without gemination (LH). This suggests that the forms with word-final LH are not at all uncommon in words like 'puppy'. This raises a new question of why word-final LH is sometimes preferred to word-final HH. If reranking of the constraint ProsForm happens, then we need to know what triggers this reranking.

6.5 Summary

In section 6.1, we started with two questions: why do geminates appear in loan-words, and why are they extremely common therein? In answer to the first question, this section sketched two approaches, an input-oriented approach and an output-oriented one. A basic claim of the input-oriented approach is that closed syllable

25 The observation that the failure of [r] to geminate is 'in sharp contrast with fricatives' is first made by Katayama (1998: 123), who claims that its non-gemination results phonologically from its markedness in the native lexicon and phonetically from its phonetic quality of tap, which is short by definition.

26 In Ito and Mester (1995: 187), different degrees of nativization common in loanwords are accounted for by the reranking of faithfulness constraints. They claim that the more the form is nativized, the lower the faithfulness constraints are ranked as against markedness constraints. The different rankings between (30a) and (30b) reflect this claim.

structures in the input source words trigger gemination, while the output-oriented approach proposes that the interaction of various constraints motivated in native phonology essentially determines gemination in loanwords. In both analyses, gemination in loanwords is motivated to achieve a certain prosodic structure, which is similar to the motivation found in mimetic vocabulary.

The answer to the question of why geminates are extremely common in loanwords can be found in the different motivations we find for gemination between loanwords and the other three lexical strata, where geminates appear either to remedy the phonotactic structure, to intensify the expression, or to show the integrity of compounding. In the native, SJ and mimetic vocabularies, the input forms are fixed and there is little chance for gemination to occur. In loanwords, in contrast, the abundance of geminates results from the structure of source forms in the input-oriented approach, and from the characteristics of loanwords in the output-oriented approach. Kubozono, Ito, and Mester (2008: 961) note that "loanwords are more faithful to the input than are native words and are free from markedness constraints to which native words are sensitive." Loanword inputs are not rigid alignments of segments, as are those input to the other lexical strata. They are more flexible in their phonological interpretation, which can give loanwords more latitude for gemination to occur.

Acknowledgements

I would like to express my gratitude to Timothy Vance, Haruo Kubozono and Shigeto Kawahara. Their invaluable comments helped me to improve my arguments and clarify my presentation. All errors that remain are my own.

References

Aoki, P. Ken. 1981. *Gemination in Japanese.* Seattle: University of Washington dissertation.

Arai, Masako and Itsue Kawagoe. 1998. Eigo no onsetsu-gata to sokuon chikaku: Nansensugo ni yoru chikaku tesuto no hōkoku [Consonant gemination and syllable types in English: A perceptual test of nonsense words]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 2(3). 87–92.

Arisaka, Hideyo. 1940. *On'inron* [Phonology]. Tokyo: Sanseido.

Best, Catherine T., Gerald W. McRoberts and Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 14. 345–360.

Bloch, Bernard. 1946. Studies in Colloquial Japanese I: Inflection. *Journal of the American Oriental Society* 66. 97–109. Reprinted in Roy Andrew Miller (ed.) 1970. *Bernard Bloch on Japanese*, 1–24. New Haven: Yale University Press.

Cho, Y. 1989. On the principle of similarity in consonantal assimilation. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 25-I*, 14–26.

Davis, Stuart. 2011. Geminates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 873–897. Malden, MA and Oxford: Wiley-Blackwell.

Davis, Stuart and Natsuko Tsujimura. 1991. An autosegmental account of Japanese verbal conjugation. *Journal of Japanese Linguistics* 13. 17–144.

Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford & Tokyo: CSLI and Kurosio.

Hashimoto, Kaoru. 1993. The relationship between pitch accent and vowel devoicing in Japanese. University of Toronto MA thesis.

Hashimoto, Shinkichi. 1950. *Kokugo on'in no kenkyū* [Studies on Japanese phonology]. Tokyo: Iwanami.

Hattori, Shiro. 1960. *Gengogaku no hōhō* [Methodologies in linguistics]. Tokyo: Iwanami.

Hirayama, Manami. 2008. Asymmetries in voiced stop gemination. In Haruo Kubozono (ed.), *Asymmetries in phonology: An East-Asian perspective*, 77–92. Tokyo: Kurosio.

Hirozane, Yoshito. 1992. Perception by Japanese speakers of some English sounds as the Japanese choked sound /Q/. *The Bulletin* (The Phonetic Society of Japan) 201. 15–19.

Hizume, Shuji. 2003. Seidaku bunka to sokuon/hatsuon [The establishment of sei-daku opposition and moraic obstruents/nasals]. *Kokugogaku* 213(2). 96–108.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Ito, Junko. 1986. *Syllable theory in prosodic phonology*. Amherst, MA: University of Massachusetts dissertation. (Garland Press, 1988).

Ito, Junko. 1990. Prosodic minimality in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II*, 213–239.

Ito, Junko and R. Armin Mester. 1993. Licensed segments and safe paths. *Canadian Journal of Linguistics* 38(2). 197–213.

Ito, Junko and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers in Linguistics 18), 181–209.

Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takeshi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin and New York: Mouton de Gruyter.

Ito, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko, Yoshihisa Kitagawa and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Jōo, Hakutarō. 1977. Gendai nihongo no on'in [Phonology of modern Japanese]. In Susumu Ōno and Takesi Sibata (eds.), *Iwanami kōza nihongo 5: On'in* [Iwanami Japanese series 5: Phonology], 107–127. Tokyo: Iwanami.

Kang, Yoonyung. 2011. Loanword phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 2258–2282. Malden, MA and Oxford: Wiley-Blackwell.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology*. Santa Cruz, California: University of California dissertation.

Kawagoe, Itsue. 1995. Shakuyōgo ni miru sokuonka to rizumu shōtotsu [Gemination and rhythm clash in loanwords]. *Gengo Kenkyu* 108. 46–73.

Kawahara, Shigeto. 2011. Japanese loanword devoicing revisited: A rating study. *Natural Language and Linguistic Theory* 29. 705–723.
Kawahara, Shigeto. 2012. Gairaigo yūsei sokuon no museika: Rironteki kōken [Japanese loanword devoicing: A review]. *On'in Kenkyū* [Phonological Studies] 15. 93–104.
Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kitahara, Mafuyu. 1997. On'inron to bunpō: Shakuyōgo no sokuon to akusento no bunseki o tsūjite [Phonology and grammar: On the analysis of geminates and accentuation in loanwords]. In Onsei Bunpō Kenkyūkai (ed.), *Bunpō to onsei* [Grammar and speech], 213–231. Tokyo: Kurosio.
Koizumi, Tamotsu. 1978. *Nihongo no seishohō* [Orthography of Japanese]. Tokyo: Taishukan.
Komatsu, Hideo. 1981. *Nihongo no sekai 7: Nihongo no on'in* [The world of Japanese 7: Japanese phonology]. Tokyo: Chūōkōronsha.
Kubozono, Haruo. 1999. *Nihongo no onsei* [The sounds of Japanese]. Tokyo: Iwanami.
Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In Caroline Féry and Ruben van de Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.
Kubozono, Haruo. 2006. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kubozono, Haruo, Junko Ito and Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. *Proceedings of the 18th International Congress of Linguists*, CD-ROM version, 953–973.
Kubozono, Haruo, Hajime Takeyasu and Mikio Giriko. 2013. On the positional assymmetry of consonant gemination in Japanese loanwords. *Journal of East Asian Linguistics* 22(4). 339–371.
Kunihiro, Tetsuya. 1963. Gairaigo hyōki ni tsuite: Nichiei on'in taikei no hikaku [On the transcription of loanwords: Comparative study of Japanese and English phonological systems]. In Tamotsu Koizumi, Shigeru Iwai and Tatsuo Nishitani (eds.), *Nichiei ryōgo no hikaku kenkyū jissen kiroku* [Reports on comparative study of Japanese and English], 27–48. Tokyo: Taishukan.
Kurisu, Kazutaka. 2000. Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9. 147–185.
Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. Cambridge, MA: MIT dissertation. (Garland Press, 1979).
Kuroda, S.-Y. 1967. Sokuon oyobi hatsuon ni tsuite [On moraic obstruents and nasals]. *Gengo Kenkyu* 50. 85–99.
Labrune, Laurence. 2012. *The Phonology of Japanese.* Oxford: Oxford University Press.
Lovins, Julie B. 1973. *Loanwords and the phonological structure of Japanese.* Chicago: University of Chicago dissertation. (Indiana University Linguistics Club, 1975).
Makihara, Hideo. 1998. Sino-Japanese compound formation and position-sensitive faithfulness constraints. Ms., University of Washington.
Martin, Samuel E. 1952. *Morphophonemics of standard colloquial Japanese.* Supplement to Language, Language Dissertation No. 47. Linguistic Society of America.
Maruta, Kōji. 1999. Eigo shakuyōgo ni okeru jūshiinka ni tsuite no ichi kōsatsu [A study on gemination in loanwords from English]. Nagoya University MA thesis.
Matsui, Michinao F. 2012. Shakuyōgo ni okeru sokuon seiki no yokusei yōin [An inhibitory factor of occurrence of voiceless geminate consonants in Japanese loanwords]. *Theoretical and Applied Linguistics at Kobe Shoin* 15. 49–101.

McCarthy, John J. and Alan S. Prince. 1986. Prosodic morphology. Ms. University of Massachusetts at Amherst.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Nasu, Akio. 1996. Niji kango ni okeru sokuonka genshō: Saitekisei riron ni yoru bunseki [Gemination in Sino-Japanese words: An optimality-theoretic analysis]. *Onsei Gakkai Kaihō* [The Bulletin, The Phonetic Society of Japan] 213. 27–40.
Nasu, Akio. 2007. Onomatope no gomatsu sokuon [The word-final moraic obstruent in Japanese mimetics]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 11(1). 47–57.
Nasu, Akio. 2008. Phonological markedness and asymmetries in mimetics. In Haruo Kubozono (ed.), *Asymmetries in phonology: An East-Asian perspective*, 49–76. Tokyo: Kurosio.
Nihon Hōsō Kyōkai (NHK) (ed.). 1987. *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary]. Tokyo: Nihon Hōsō Shuppan Kyōkai.
Nishimura, Kohei. 2003. Lyman's Law in loanwords. Nagoya University MA Thesis.
Ohso, Mieko. 1971. A phonological study of some English loan words in Japanese. Ohio State University MA thesis.
Ohye, Saburo. 1967. Gairaigochū no sokuon ni kansuru ichi kōsatsu [The mora phoneme /Q/ in English loanwords in Japanese]. *Onsei no Kenkyū* [The Study of Sounds, The Phonetic Society of Japan] 13. 111–121.
Okumura, Mitsuo. 1977. On'in no hensen [Phonological changes]. In Susumu Ōno and Sibata Takesi (eds.), *Iwanami kōza nihongo 5: On'in* [Iwanami Japanese series 5: Phonology], 221–252. Tokyo: Iwanami.
Ono, Kōji. 1991. Gairaigo to shite no eigo no sokuonka ni tsuite [On gemination of loanwords from English]. *Gengo Kenkyu* 100. 67–87.
Paradis, Carole and Jean-Francois Prunet (eds.). 1991. *The special status of coronals: Internal and external evidence* (Phonetics and Phonology 2). San Diego: Academic Press.
Peperkamp, Sharon, Inga Vendelin and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25. 129–164.
Poser, William John. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.
Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Ms., Rutgers University Center for Cognitive Science. (Blackwell, 2004).
Saito, Michiaki. 1992. *Gendai nihongo no gokōseironteki kenkyū* [A morphophonological study of modern Japanese]. Tokyo: Hituzi Syobo.
Shirai, Setsuko. 1999. Gemination in loans from English to Japanese. University of Washington MA thesis.
Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.
Suzuki, Hisami. 1995. Minimal words in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 31*, 448–463.
Takagi, Naoyuki and Virginia Mann. 1994. A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22. 343–356.
Takayama, Tomoaki. 1995. Sokuon ni yoru fukugō to takuritsu [Compounding and prominence by gemination]. *Kokugogaku* 182. 15–27.
Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Tamamura, Fumio. 1989. Gokei [Word forms], In Fumio Tamamura (ed.), *Kōza nihongo to nihongo kyōiku 6: Nihongo no goi, imi (jō)* [Japanese and Japanese teaching 6: Japanese words, meanings 1], 23–51. Tokyo: Meiji Shoin.

Tateishi, Koichi. 1989. Theoretical implications of the Japanese musicians' language. *Proceedings of West Coast Conference on Formal Linguistics 8*, 384–98.
Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13. 209–235.
Tsuchida, Ayako. 1995. English loans in Japanese: Constraints in loanword phonology. *Working Papers of the Cornell Phonetics Laboratory* 10. 145–164.
Tsujimura, Natsuko and Stuart Davis. 1988. The morphophonemics of Japanese verbal conjugation: An autosegmental account. *Proceedings of the Eastern States Conference on Linguistics 5*, Columbus, OH: Ohio State University, 488–499.
Ura, Hiroyuki. 1995. A metrical and optimality-theoretical approach to gemination and epenthesis in Japanese loanword phonology. *Osaka University Papers in English Linguistics* 2. 153–196.
Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.
Vance Timothy J. 2002. Semantic bifurcation in Japanese compound verbs. In Noriko M. Akatsuka and Susan Strauss (eds.), *Japanese/Korean linguistics 10*, 365–377. Stanford: CSLI.
Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.
Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Waida, Toshiko. 1984. English and Japanese onomatopoeic structures. *Joshidai Bungaku (Gaikoku Bungakuhen)* 36. 55–79. Osaka Women's University.
Yamaguchi, Yoshinori. 1989. Nihongo no rekishi: On'in [History of Japanese: Phonology]. In Takeshi Kamei, Rokurō Kōno and Ei'ichi Chino (eds.), *Gengogaku daijiten 2: Sekai gengohen 2* [The Sanseido encyclopedia of linguistics 2: Languages of the world, Part 2], 1634–1647. Tokyo: Sanseido.
Yanagida, Seiji. 2003. On'inshi [History of Japanese phonology]. In Zendo Uwano (ed.), *Asakura nihongo kōza 3: Onsei on'in* [Asakura Japanese series 3: Phonetics and phonology], 43–60. Tokyo: Asakura Shoten.

Gábor Pintér

3 The emergence of new consonant contrasts

1 Introduction

One of the better known features of Japanese segmental phonology is its restrictive phonotactic system. Unlike most European languages, Japanese has numerous co-occurrence restrictions on the possible consonant-vowel combinations. For example, both the voiceless alveolar fricative [s] and the high vowel [i] are regular segments in the Japanese sound system, yet they cannot be combined. Examples from loanword phonology in (1) and verb paradigms in (2) show that the absence of [si] sequences from Japanese is not accidental; they are systematically avoided in the language. In this chapter, unlike other chapters in the same volume, the Hebon-shiki Romanization is used to represent Japanese words since the Kunrei-shiki Romanization is not always useful or accurate in describing sounds whose phonemic status is still controversial.[1] In addition, [u] rather than [ɯ] is used as the phonetic representation of /u/ since it is not clear when exactly /u/ lost lip rounding in the course of the history.

(1) Adaptation of [si] as [ʃi] in loanwords

'seat' → [ʃiː]to	'basic' → bee[ʃi]kku
'seafood' → [ʃiː]fuudo	'Lucy' → ruu[ʃiː]

(2) The [s]-[ʃ] alternation in verb conjugation

hana[s]-*u*	speak-NON-PAST
hana[s]-*anai*	speak-NEG
hana[ʃ]-*imasu*	speak-POLITE
hana[s]-*eba*	speak-COND

The two most common and straightforward methods for describing this kind of phonotactic restriction are invoking a constraint against the unacceptable CV sequences (e.g., *SI), or by formulating processes that change the illicit combinations of segments into acceptable ones (e.g., /si/ → [ʃi]). While these techniques have become commonplace technical tools in the arsenal of practicing phonologists, they do not offer much insight into the intricate hierarchy that exists among the CV constraints. The latest developments in the Japanese sound system show evidence for a general tendency towards a less restrictive phonotactic system. Some of the constraints that used to be active before the turn of the 20th century are now inactive. For example, the constraint on *[ti] sequences is preserved only in older lexical

1 For example, [ti] and [tʃi] are written as *ti* and *chi*. In the Kunrei style, both syllables may be transcribed as *ti*. Long vowels are represented as double vowels although this is not a common practice in the Hebon-shiki Romanization: e.g., *shiito* 'seat'.

borrowings (e.g., 'team' → *chiimu*); newer loanwords can contain [ti] (e.g., 'party' → *paatii*). Some constraints are in the process of phasing out. For instance, Kubozono (Ch. 8, this volume) reports an increase in loanwords containing innovative [tu] syllables instead of the conservative [tsu] (e.g., 'Bantu' → *bantuu~bantsuu*). While opinions may diverge about the phonemic status of [tu] in contemporary Japanese, the tendency to integrate this syllable into the language is undeniable. In spite of the general loosening of phonotactic constraints, some of the CV restrictions are adamant, not showing the slightest signs for a change. For example, there is no evidence hinting that [hu] or [çe] could emerge as new syllables in Japanese. These sequences are unexceptionally avoided in Japanese. The differences in the strength of the CV restrictions together with the speaker's intuition about the grammaticality of novel forms are not accidental. New syllables do not just emerge randomly in the sound system. The primary goal of this research is to reveal the hierarchical relations among phonotactic restrictions and provide explanations for them.

Before going into details a note about the relation between gaps and contrasts is due here. The phonotactic gaps in the syllabary are interpreted as the lack of consonantal contrasts in given vocalic environments. For example, the absence of [si] is viewed as the lack of the /s/ ↔ /ʃ/ opposition in the context of /i/. Similarly, the emergence of new syllables is described as the extension of consonantal contrasts to new vocalic environments. For example, the birth of syllable /ti/ in Japanese is interpreted as the extension of the /t/ ↔ /tʃ/ contrast to the environment of /i/. The members of the contrasts are selected based on evidence from loanword phonology (see (1)), morphological kinship (see (2)) as well as misperception and mispronunciation in second language learning.

Having looked at the Japanese syllabary, it is immediately clear that consonant contrasts regularly avoid the vocalic environments of /i/, /u/, and /e/ (e.g., /si/ = /ʃi/, /hu/ = /ɸu/, /he/ = /çe/). The central tenet of this study is that these common phonotactic patterns are results of recurrent diachronic changes. The similarities across sound changes, even across centuries, are due to shared motivations originating in universal principles of perceptual and articulatory phonetics (Ohala 1974, 1981; Blevins 2004). The current system of phonotactic restrictions is best understood through historical investigations of the emerging contrasts.

The historical development of Japanese consonants displays several recurring patterns. First, most innovative consonant phonemes in Modern Japanese go back to an allophonic status. Second, the allophones are typically results of articulatorily motivated changes. Third, loanwords play a crucial role in the phonologization of allophones and the distributional expansion of new contrasts. Fourth, the emergence of innovative consonants in new vocalic environments is subject to perceptual constraints. Fifth, interestingly, the perceptual constraints veto exactly those vocalic environments that gave birth to allophones in the first place. These vocalic environments typically involve high vowels. For example, the voiceless alveolar affricate [tʃ] first appears as an allophone of /t/ in a historical [ti] > [tʃi] assibilation. While

the high front vowel [i] is responsible for triggering assibilation, it is also the most resistant environment to the emerging /t/ ↔ /tʃ/ contrast in Modern Japanese.

This study reveals how recurrent patterns in the history of Japanese consonant contrasts were shaped by regular articulatory and perceptual forces combined with the external influence from loanwords. Section 2 gives a brief overview of Japanese consonants and their uneven distribution over vocalic environments. Section 3 then compares how Japanese consonant-vowel restrictions might be analyzed in various phonological frameworks. This section also introduces the combination of perceptual and acoustic forces which explains common phonotactic asymmetries. Section 4 enumerates and explains how phonological contrasts in the series of voiceless fricatives, stops and affricates evolved through time in Japanese. Section 5 goes on to examine the implications of Japanese phonotactic changes for theories of phonological representation. The last section concludes that the overall development of Japanese consonants follows a general historical pattern in which sound changes are perception-oriented, while the seeds for change have articulatory origins (Hyman 1976: 416). Besides supporting this general concept, the article presents several episodes from the history of Japanese consonants illustrating how uneven pressure from loanwords and telescopic effects of sound changes can also lead to phonotactic asymmetries.

2 Phonotactic gaps in modern Japanese

The inventory shown in (3) summarizes the consonantal segments of Modern Japanese as described in standard textbooks (Shibatani 1990: 159; Tsujimura 2007: 15; Labrune 2012: 59). Those sounds whose phonemic status is not unanimously acknowledged are surrounded by parentheses. These problematic sounds typically fall into two groups. They are either consonants whose occurrence is restricted to recent loanwords (e.g., 'fan' → /ɸan/) or surface allophones appearing as outputs of assimilations (e.g., /hj/ → [ç]; /sj/ → [ʃ]). The phonological treatment of these sounds is a somewhat complicated topic, which is discussed in detail in section 5.

(3) Japanese consonant phoneme inventory

	labial	alveolar	alveo-palatal	palatal	velar	uvular / glottal
stop	p b	t d			k g	
fricative	(ɸ)	s z	(ʃ)(ʒ)	(ç)		h
affricate		(ts)(dz)	(tʃ)(dʒ)			
approximant		r		j	w	
nasal	m	n	(ɲ)			N

The way Japanese consonants can combine reveals a fairly restrictive phonotactic system. For example, Japanese does not allow for onset consonant clusters. Onset clusters from foreign words are broken up by intrusive vowels (e.g., 'strict' → *sutorikuto*). The coda can only accommodate non-contrastive nasal stops or obstruents that form the first part of a geminate (e.g., [-k.k-], [-p.p-] but *[-p.t-], *[-k.t-], *[-t#], *[-s#]).

The most intriguing feature of Japanese phonotactics is that consonants cannot combine freely with all the possible vowels in CV sequences. The co-occurrence restriction between consonants and consecutive vowels deserves special attention because in most phonological frameworks, consonant-vowel sequences are analyzed into separate syllable constituents with no phonotactic interactions assumed between them.[2] Vowels, as members of the nucleus, can freely combine with onset consonants in numerous – typically Western – languages. Interestingly, it is exactly this combinatorial freedom between onsets and rhymes, in sharp contrast with within-constituent restrictions, that led phonologists settle with a right branching structure for the syllable (Steriade 1988).

Because of the numerous ongoing phonotactic changes in Japanese, gaps in the syllabary represent a moving target. When discussing Modern Japanese, it is necessary to distinguish between conservative and innovative varieties of the language (Vance 1987: 17). The conservative variety is associated with an earlier language state that is yet to be affected by modern loanwords. This variety displays a large number of CV restrictions, as shown in (4).

(4) Some syllables in the conservative variety (surface forms)

	s	ʃ	h	ç	ɸ	t	tʃ	ts
a	sa	ʃa	ha	ça		ta	tʃa	
i		ʃi		çi			tʃi	
u	su	ʃu		çu	ɸu		tʃu	tsu
e	se		he			te	tʃe	
o	so	ʃo	ho	ço		to	tʃo	

The Meiji Restoration in 1868 not only opened up Japan politically, but also made the language adaptive to foreign influence. The volume of lexical adaptations has been so massive that it led to the birth of a new stratum in the lexicon (Ito and Mester 1995; Nasu, this volume; Ito and Mester, this volume; Kubozono, Ch. 8, this volume). In response to the pressure from loanwords, the phonotactic system has

2 But see Kawasaki (1982) or Janson (1986) on onset-nucleus restrictions.

been showing a steady increase in the number of allowed syllable types. The innovative variety is associated with a language state that accommodates (almost) all logically possible CV combinations, as shown in (5).

(5) Some syllables in the innovative variety (surface forms)

	s	ʃ	h	ç	ɸ	t	tʃ	ts
a	sa	ʃa	ha	ça	ɸa	ta	tʃa	tsa
i	si	ʃi		çi	ɸi	ti	tʃi	tsi
u	su	ʃu		çu	ɸu	tu	tʃu	tsu
e	se	ʃe	he	çe	ɸe	te	tʃe	tse
o	so	ʃo	ho	ço	ɸo	to	tʃo	tso

The actual set of syllables most speakers use varies between the idealized extremes of the conservative and innovative varieties. For example, an average speaker would pronounce [ʃe] without any hesitation, but almost no one would use [si] in an everyday discourse.

3 Theories

Interestingly, the topic of CV restrictions does not enjoy very much attention in the literature of Japanese phonology. Although the phonotactic restrictions are widely acknowledged, they seldom serve as the subject matter for phonological studies. The following paragraphs enumerate some delightful exceptions to this generalization.

3.1 The theory of "sukima" and "akima"

One of the few approaches to Japanese CV restrictions in the literature is presented by Hattori Shiro (Hattori 1960). His description of the phonotactic gaps relies heavily on assimilatory processes such as /si/ → [ʃi]. He points out that those surface forms that can stand at the input side of assimilations (e.g., [si] or [hu]) form a special type of gap, termed "sukima" (Hattori 1960: 289, 317). These gaps are typically difficult to fill because assimilations remove them from the surface. These systematic gaps are distinguished from another type of phonotactic restrictions called "akima".[3] *Akima* refers to the absence of logically possible and phonotactically legal combinations

3 The word *sukima* translates to English as 'gap', *akima* as 'vacancy'.

of phonemes. Cases of *akima* can be roughly interpreted as accidental gaps in the phonological system. For example, in Hattori's interpretation, the absence of /ʃe/ is an *akima* type of gap because there is no assimilation that targets /ʃe/ sequences. Both /ʃ/ and /e/ are valid phonemes; the lack of their combination is accidental. Hattori notes that these accidental gaps are easier to fill than those that are systematically eliminated by assimilations.

The distinction between cases of *sukima* and *akima* boils down to the presence versus absence of assimilatory processes, and consequently to the phonemic status of the consonants involved. If the surface consonant is treated as an allophone, that is, as an output of an assimilatory process, then the gap it leaves behind is a *sukima*. The sibilant [ʃ] in [ʃi] does not go back to an underlying /ʃ/ phoneme but results from a /si/ → [ʃi] assimilation. Thus, [si] is a *sukima*. If we presume phonemic status for the fricative in [ʃi], then no assimilation takes place (i.e., /ʃi/ → [ʃi]), and the resulting gap of [si] is an *akima*. Without assimilatory rules, the absence of [si] on the surface is analyzed as the accidental absence of /s/ plus /i/ phoneme sequences.

The ongoing process of phonologization in Japanese consonants can be viewed as the fossilization – and decline – of assimilatory rules. When the assimilation becomes inactive, the output consonant is reanalyzed as an independent phoneme. That is, systematic gaps are reanalyzed as accidental ones. Hattori takes the case of [tʃi] in Modern Japanese as an example of such fossilization. In his analysis, the /ti/ → [tʃi] assimilation is not active anymore.[4] The surface [tʃi] form comes from underlying /tʃi/ without the application of allophonic rules (Hattori 1960: 289). Thus the absence of [ti] is an accidental gap. The fossilization of the *ti*-assimilation is also documented in the Japanese vocabulary, as [ti] is replaced by [tʃi] in older loanwords (e.g., 'plastic' → *purasuchikku*), whereas it is preserved as [ti] in new loans (e.g., 'stick' → *sutikku*).

The most important contribution of Hattori's work was to make a distinction between different types of CV restrictions, and to involve assimilatory rules in the explanation of rigid phonotactic constraints. His approach, however, leaves several questions open. First, there is an apparent circularity in the discussion of assimilatory rules and the difficulty of filling phonotactic gaps. On one hand, the difficulty of adapting certain syllable types is explained by the presence of assimilatory rules. On the other hand, the question of whether assimilatory rules are present or not is decided with reference to the difficulties to fill those gaps that the assimilation creates. Regardless of the circularity, Hattori's theory of gaps may be valid, but justification for it has to be found somewhere else.

The second problem, closely related to the first one, is the lack of criteria in determining whether an allophonic rule is fossilized or not. In other words, how can it be tested if a surface form is just a positional allophone or has already earned

4 Hattori uses the symbol /c/ for the voiceless affricate (here /tʃ/). See Nishida (2010) for a direct critique of Hattori's analysis using underlying affricates.

a phonemic status? Does [tsu] come from an underlying /tsu/ or is it derived from /tu/? The criterion remains unclear. Third, hierarchical relationships between gaps of the same type, that is within *sukima* or *akima* types, are not considered. Some of the systematic gaps are filled easily, (e.g., [tu]), while others are difficult (e.g., [si]).

Irrespective of all the critiques, Hattori's categorization of phonotactic gaps, together with his observations about the role of assimilatory rules, is outstanding. While his work presents an intuitively appealing description of CV phonotactics, his idea requires further elaboration. Specifically, the motivations behind the asymmetric patterns need further investigation.

3.2 Hierarchical phonotactic constraints in the lexicon

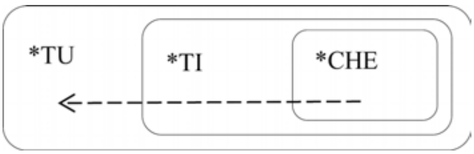
The Japanese lexicon is divided into four strata: native (or Yamato) words, Sino-Japanese words, mimetic expressions, and recent loanwords (Shibatani 1990: 140–157; Ito and Mester 1995; Labrune 2012: 13–24; Nasu, this volume; Ito and Mester, this volume; Kubozono, Ch. 8, this volume). It is not uncommon in Japanese phonological studies to formulate generalizations restricted to only certain domains of the lexicon. For example, McCawley (1968) accepts /h/ as a phoneme only within loanwords and mimetic expressions, but not in native and Sino-Japanese words.

Although connecting phonotactic constraints to certain lexical domains was not a particularly novel idea, Ito and Mester (1995) managed to address this issue from a rather illuminating angle. According to their views, the subtly defined hierarchy of phonotactic restrictions reflects the structure of the lexicon. The constraints outline a core-periphery structure in which the distance from the core symbolizes the level of integration of lexical items. Items that are more native-like obey more constraints, thus they are closer to the core of the lexicon. Less native-like items, such as loanwords, violate more phonotactic constraints; they are located on the periphery. For example, the constraint against singleton *p* (i.e., *P) demarcates the core of the lexicon. Yamato and Sino-Japanese words do not violate this constraint, whereas loanwords typically do. The loanword *sepaado* 'shepherd' has a singleton *p*, so it is on the opposite side of the constraint domain. The word *sheepaa* 'shaper' is even further away, as it violates an extra constraint that penalizes [ʃe]. Assuming that the constraints form proper subsets (Ito and Mester 1995: 830), the core-periphery relation can be plausibly demonstrated by a Venn diagram. The older loanword *sepaado* is closer to the core than the newer *sheepaa* because it has fewer violations. This model correctly predicts that there should be no loanwords in Japanese in which the singleton [p] is avoided but [ʃe] is adapted faithfully, as represented in the Venn diagram in (6).

(6) The relation of *P and *SHE constraints (Ito and Mester 1995: 830)

Although this chapter is not directly concerned with the structure of the Japanese lexicon, the concept of constraint hierarchies can be applied directly in the analysis of emerging syllables. Innovative syllables can be analyzed by extending the lexicon with new constraints at the periphery. The more recent the opposition is, the further away the corresponding constraint is from the core. For example, the diachronic order of emerging syllables of [tʃe] ≫ [ti] ≫ [tu] can be represented through concentric constraints, as shown in (7).

(7) The relation of *CHE, *TI, and *TU constraints



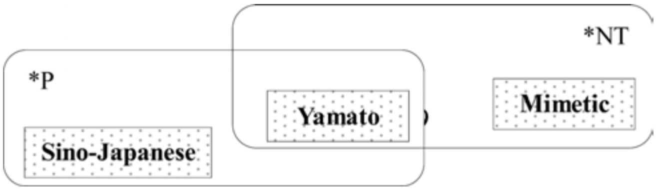
This solution, together with its variants that rely on linearly-ordered constraints (e.g., Optimality Theory) faces a few problems. First, unless the ranking of constraints is based on some sort of evidence, the theory is prone to over-generation, and loses on predictive power. Nothing prevents us from creating grammars with unrealistically ordered constraints. Furthermore, this freedom makes it difficult to make predictions. Let us invoke the constraints against [tu] and [hu] syllables in Ito and Mester's description (Ito and Mester 1995: 826). Since both constraints are obeyed in the whole lexicon, they are assigned identical scopes. In its original form the theory does not, and actually cannot, decide the ordering of the constraints, and cannot make predictions about the grammaticality of the involved syllables. By now we know that [tu] is already being integrated into the language, while [hu] is not likely to become a Japanese syllable in the foreseeable future. Although predictions are not entirely absent from Ito and Mester's theory, it leaves plenty of room for further improvements.

Another problematic feature of the constraint domain model is that the relationship between constraints can be either concentric or partially overlapping. While most of the constraints are assumed to be concentric, the relation of *P and *NT constraints,[5] as proposed by Ito and Mester (1995: 823), is an example of a non-concentric case. As it is shown in (8), mimetic words can violate *P but obey *NT;

5 *NT is a constraint against non-homorganic nasal-stop clusters such as [-mt-].

Sino-Japanese words can violate *NT but obey *P. Although allowing for non-concentric arrangement of constraints is one of the strengths of the framework, it is not obvious when constraints should be arranged this way. Moreover, it is not clear how partially overlapping constraints can be linearized along the core/periphery distinction, or along the timeline in a diachronic interpretation. Adding more constraints to the analysis aggravates these problems further, introducing unmanageable complexities to the Venn diagram model.

(8) Non-concentric constraint domains



Approaching the problem from a strictly logical point of view, it can be stated that one constraint splits the lexicon into two regions: a region in which items obey the constraint, and one in which they do not. Two constraints create four possibilities (see *P and *NT). With *n* number of constraints there are 2^n possibilities. Obviously, not all logically possible patterns are observed. Structural axioms (e.g., all constraints have to intersect with some part of the core), observational generalizations (e.g., all words must obey *[hu]), and logical relations between constraints (e.g., the special case precedes the general one) can greatly reduce the problem. Still, the number of possible combinations is very high. It is a geometrical challenge on its own to create a Venn diagram to represent relations between constraints at higher orders (e.g., Ito and Mester 1995: 834). Instead of trying to collapse all constraints into a two dimensional plane, it may be more plausible to use constraint sub-hierarchies (Padgett 2001), or a multi-dimensional constraint space in which related constraint families occupy identical dimensions (Trón and Rebrus 2001). The ranking of the constraints within a sub-hierarchy or dimension should be backed up by evidence from articulatory, perceptual or other cognitive domains. While the demarcation of constraint families seems like an extra task, it has the benefit of freeing the analysis from the burden of finding ranking arguments for cases in which the participating constraints are orthogonal. The independent ranking of constraints in different sub-hierarchies or dimensions can be interpreted as a device that describes the idiosyncratic features of languages. For example, on articulatory phonetic grounds it is reasonable to group together constraints against voiced geminate consonants (e.g., *GG > *DD > *BB), but this group of constraints is independent of the constraints that, for instance, express preference order of glide-vowel syllables (e.g., *WU >

*WO > ... etc.). Preferences for voiced geminates and glides are treated independently in languages (see Kawahara, this volume, and Kawagoe, this volume, for full discussion of geminate consonants in Japanese).

In conclusion, it can be said that Ito and Mester's study is an important milestone on the way to discovering the phonotactic hierarchy of Japanese syllables. Although the framework can formally express subtle relations among phonotactic constraints, a serious improvement could be achieved by grouping those constraints together whose universal rankings are defined by articulatory, perceptual or other independent evidence.[6]

3.3 Perceptual factors

Human speech perception provides a possible source to back up constraint-based frameworks. The current description of emerging consonantal contrasts is based on the recognition that the diachronic order of events in the consonantal system is greatly influenced by perceptual characteristics of consonants. The role of perceptual factors in phonology has a long history: it can be derived from the functional phonological principle of the Least Effort (Zipf 1949; Flemming 1995; Boersma 1998). The principle of Least Effort assumes that speakers tend to minimize articulatory effort, whereas listeners opt for minimizing perceptual confusion. The perceptual aspect of the principle ensures that linguistic signs are kept apart. By definition, perceptual forces are not to be interpreted on individual linguistic forms, but on contrasts. The perceptual salience – or perceptual fitness – of a form is defined in relation with the contrasts it is engaged in. A linguistic form is perceptually preferable if it can be sufficiently differentiated from other forms, and from silence, for that matter.

In this light, the phonotactic constraints formulated by Ito and Mester (1995) are misleading. For instance, it is fallacious to explain the absence of [hu] by a simplistic *HU constraint, even if the constraint is meant to express the poor perceptibility of [hu]. As a matter of fact, Japanese listeners *do* perceive the acoustic signals transcribed as [hu]; the problem is that they cannot tell it apart from [ɸu]. Accordingly, it is not the [hu] sequence in itself that is recognized poorly by the ears of native speakers of Japanese but rather the [hu] ↔ [ɸu] opposition.

An interesting feature of the perceptual approach is that constraints are defined over oppositions, not over individual items. As a corollary, perception does not express preferences for members of contrasts. If a contrast is neutralized, its direction is decided by non-perceptual factors. For example, until before Modern Japanese,

6 See McMahon (2000) for a similar view arguing for a finer categorization of constraints in Optimality Theory.

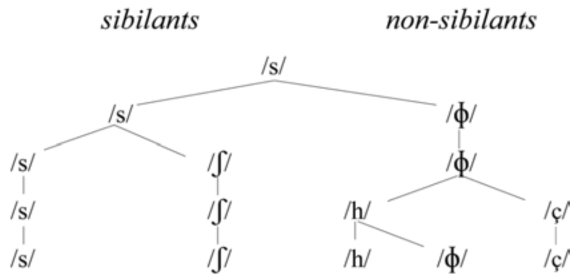
[wo] and [o] used to neutralize into [wo] in response to a strong requirement for syllable onsets in the language. After this requirement disappeared, the neutralization prefers [o] instead (Pintér 2005). Perceptual constraints in themselves cannot settle questions of directionality in neutralizations (Padgett 2001: 193).

3.3.1 Contrast dispersion

In the long run, the requirement for perceptually distinct oppositions leads to inventories that are not concentrated in a single acoustic region but spread out in the available perceptual space. The more distant the items are in perception, the less confusion they cause. The concept of maximizing perceptual contrast in inventories is also known as dispersion (Lindblom 1983; Flemming 2001; Padgett 1995). The most obvious example for dispersion is the vowel space. Vowel systems from a variety of languages demonstrate that vowels tend to occupy the vowel space in a more or less even manner (Liljencrants and Lindblom 1972; Lass 1984: 142; de Boer 2001). Not only vowels, consonants or syllables are also subject to this principle.

The evolution of the Japanese fricative system represents an illuminative example for dispersion evolving through time. Old Japanese is reconstructed with a single fricative sound which is a sibilant. Typological studies show that sibilants are the most frequent fricatives; if a language has only a single fricative, it is likely to be a sibilant (Maddieson 1984: 44, 52). The preference for sibilants can be related to their strong audible friction, which makes them perceptually substantially different from other manners of articulation. The next fricative in the history of Japanese is the bilabial non-sibilant fricative (i.e., [ɸ]). The Japanese non-sibilant/sibilant system is in accord with Maddieson's typological observation claiming this type to be the most frequent among two-fricative systems (Maddieson 1984: 14). The second most frequent two-fricative system in the typological study has two sibilants. From a perceptual point of view, both the sibilant/non-sibilant and the two-sibilant systems are better than those with only non-sibilant fricatives. Acoustic differences across the sibilant/non-sibilant border have an obvious perceptual benefit compared to within-category distinctions. Because of their greater acoustic strength, a pair of sibilants is still better than two non-sibilant ones. The tendency to avoid two non-sibilant fricatives manifests itself in the next historical step in Japanese which features two sibilant and a non-sibilant fricative (i.e., /s ʃ ɸ/). This system is the most frequent configuration in languages with a three-way fricative distinction (Maddieson 1984: 54). The last steps in Japanese involve additions of two non-sibilant fricatives /ç/ and /ɸ/. The evolution of the fricative system is summarized in (9).

(9) The evolution of the fricative system in Japanese



Even if the overview radically simplifies both the historical and phonetic details, it is clear that the Japanese fricative system evolved through stages that meet the requirement of dispersing contrast. The data also demonstrates that even though historical changes are believed to have articulatory seeds, the outcome displays the influence of perceptual factors.

3.3.2 Perceptual licensing

When a segmental contrast appears in a language or when one disappears, it does not happen overnight. Changes take place in small proportions affecting different linguistic contexts unequally. An important source for these disparities is related to the fact that the same consonantal opposition is not equally salient in all the positions in which it occurs. The phonetic context can greatly influence the salience of perceptual cues. Those contexts which provide a greater number, or more robust cues, are better licensers of perceptual contrasts. In a poorly licensed environment perceptual cues are either reduced or eliminated. The typologically frequent word-final devoicing of obstruents is an example of poor perceptual licensing. Although the voicing contrast can utilize several acoustic cues (Lisker 1986), the most important ones, realized in CV transitions, are absent in word-final positions. The smaller number of cues leads to diminished perceptual licensing capacity, which explains the popularity of this site for voice neutralization (Steriade 2001; Blevins 2004: 94–95).

Formant transitions in CV sequences provide another prominent example for perceptual licensing. The opposition between such syllables as [wa] versus [a], [we] versus [e] can be analyzed as a [w] ↔ Ø contrast in different vocalic environments. The formant transitions that lead from the labio-velar glide to a subsequent vowel are responsible for cuing the contrast. Since the formant structure of [w] shows close resemblance to that of [u], the transitional cues are minimal, if they exist at all, in the context of [u] (Padgett 2001: 193). This poor capacity of perceptual licensing

explains the cross-linguistically well-documented tendency to avoid /wu/ ↔ /u/ contrasts.

Recognizing the differences in the perceptual licensing potential of different phonetic contexts allows us to apply the principle of dispersion to sub-inventorial levels of description. Dispersion theory predicts that the particular order in which an emerging contrast occupies new environments reflects the perceptual licensing capacity of the contexts. Due to the subtleties of language change, it is possible that a contrast appears earlier in an environment that is perceptually more challenging than an unutilized one. In this case, the theory predicts that the discontinuity in the distribution of the contrast is accidental, and the filling of the gap should pose no problem for the speakers. Dispersion also predicts that disappearing contrasts hold on the longest to those contexts in which they are cued more robustly. For example, the decline of [w] ↔ Ø contrasts in Japanese syllables left a single /wa/ ↔ /a/ opposition in the system before modern loanwords reversed the process. Interestingly, the order of decline coincides with the perceptual licensing capacities of the vocalic contexts (Pintér 2005).

3.4 Articulatory factors

The articulatory aspect of the Least Effort principle requires speakers to minimize their efforts when producing speech. This type of laziness is not specific to speech; it is present in any muscle activity (Lindblom 1983). It explains why movements, articulatory and otherwise, tend to get shorter, why ballistic movements tend to be preferred over precisely controlled ones, and why subtle timing of gestures often results in loss of coordination. These articulatory forces represent the major sources in creating phonetic variability as well as the allophonic bases for phonologization.

Assimilation is a typical example for articulatory economy. By definition assimilation refers to the processes in which (usually neighboring) segments become more alike. For example, in casual English, the alveolar stop /t/ and the palatal glide /j/ can assimilate into an affricate in such frequent phrases as ‘don’t you’ [‘doʊntʃə] (i.e., /tj/ → /tʃ/; Crystal 1980: 31; Clark and Yallop 1990: 122). Since the biological buildup of humans, as well as their articulatory organs, are by and large identical, similar articulatorily-motivated changes are expected to occur in unrelated languages in a parallel fashion.

At this point, it is worth having a closer look at the division of labor between perceptual and articulatory factors. Assimilatory changes demonstrate articulatory forces can influence pronunciation directly, producing new sounds in the sound system. Perceptual factors work in a radically different way. Perceptual forces are passive; they manifest themselves indirectly by filtering out sub-optimal forms.

Perception by itself cannot create or change oppositions no matter how preferable the change would be on perceptual grounds.[7]

3.5 Other factors

While articulatory and perceptual principles are frequently used to analyze historical changes, the role of loanwords as the catalyst of change tends to be ignored in phonological analyses (see Kubozono, Ch. 8, this volume, for related issues). Considering the history of Japanese, it can be said that changes caused by loanwords in the consonant system are comparable to the changes caused by internal motivations. Both Chinese and Western borrowings made significant additions to the lexicon of Japanese. Chinese had a strong influence on syllable structure (Frellesvig 2010: 184); Western languages contributed greatly to the elimination of constraints on consonant-vowel sequences. Some of the phonotactic peculiarities of the Japanese syllabary are difficult to explain on phonetic grounds, because they are just phonological, inventorial footprints of donor languages. For example, the near absence of words with [tsa] ('tsar'$_{\text{RUSSIAN}}$ → *tsaaru*; 'Mozart'$_{\text{GERMAN}}$ → *mootsaruto*) in contemporary Japanese can simply be attributed to the fact that affricate [ts] is absent from English onsets, which is the main donor language of loanwords in Modern Japanese (Irwin 2011: 25; Kubozono, Ch. 8, this volume). The number of lexical borrowings from Italian, German and Russian containing [ts] is small, but no articulatory or perceptual difficulties are reported in connection with [ts] in these words. The marginal presence of /ta/ versus /tsa/ contrast in contemporary Japanese is not perceptual in nature. It is just a result of coincidence of non-linguistic factors that there is only negligible external motivation for establishing /tsa/ syllables in Japanese.

Since pressure from external and internal sources has an acknowledged role in the analysis, it is misleading and fallacious to require perceptual or articulatory explanations for linguistics forms whose absence is attributable to lack of tendencies to create them in the first place. Perceptual constraints can predict the *possible* consonant-vowel sequences in the language, but they cannot explain which sequences are actually realized. Historical investigations combined with phonetic experiments may prove that gaps that are believed to be systematic are in fact accidental.

4 Historical investigations

The following subsections explain how certain consonant phonemes emerged in Japanese with special attention to how articulatory and perceptual forces are involved in the process. Due to limitations on space, only the voiceless fricative and the voice-

7 The indirect nature of perceptual mechanisms can lead to false interpretations about its role in language change. See Ohala (1981) for further clarifications.

less stop/affricate systems are discussed. Also, for sake of brevity, the descriptions of historical events are kept simple. For more elaborate historical studies, the reader is advised to consult Miller (1967), Martin (1987), Unger (1993) and Frellesvig (2010) as well as the chapters in the History Volume.[8]

4.1 Sibilant fricatives

There are five fricative consonants in contemporary Japanese: [s], [ʃ], [h], [ç], and [ɸ]. Dividing them into sibilants and non-sibilants has not only phonetic but also diachronic motivations. The sibilant fricatives go back to a single sibilant fricative in Old Japanese; the non-sibilant ones are the distant descendants of bilabial stops.

4.1.1 The birth of sibilant allophones

There is a broad consensus among historical linguists that Old Japanese (8th century) had a single fricative element, which was a sibilant, perhaps an affricate sound (Hashimoto 1950; Lange 1973; Martin 1987; Unger 1993). While there is no disagreement about the existence of a sibilant-like phoneme in Old Japanese, the debate concerning its exact phonetic value is still unsettled. The following table summarizes the most prominent hypotheses concerning the phonetic value of /s/ in Old Japanese.

(10) Hypothetical phonetic values of Old Japanese sibilants

Modern pronunciation		[sa]	[ʃi]	[su]	[se]	[so]	*notes*
reconstructions	Arisaka* (1944: 143)	tsa	si	tsu	se	tso	**cited in Miller (1967: 192)*
	Lange (1970: 134)	Sa	Si	Su	Se	So	*S : [s] or [ʃ] but not [ʦ]*
	Unger (1993: 22)	sa	si	su	sye	so	
	Mabuchi (1971: 96)	ʃa	ʃi	ʃu	ʃe	ʃo	
	Frellesvig* (2010: 36)	^{t}sa	ʃi	^{t}su~su	ʃe	^{t}so	**following Kobayashi (1981)*

It is widely accepted that by Late-Old Japanese, /s/ had an alveo-palatal or post-alveolar allophone before front vowels. Some of the hypotheses claim that the sibilant system was already allophonic by the time of Old Japanese (e.g, Frellesvig 2010). Some others (e.g., Mabuchi 1971; Unger 1993) assume that /s/ in Old Japanese had a uniform pronunciation which later split into two allophones, as shown in (11).

8 For Japanese descriptions about the history of the language, the reader can consult Mabuchi (1971), Watanabe (1997) and Nishida (2001), among others.

(11) Allophones in voiceless sibilant fricatives

	Late-Old Jp.	
	s	ʃ
a	sa	
i		ʃi
u	su	
e		ʃe
o	so	

Logically speaking, there are two possible paths to reach the allophonic distribution from a uniform one. In the traditional view, the [s]-[ʃ] allophony is a result of typologically trivial, context-sensitive assimilation of [si] > [ʃi] and [se] > [ʃe]. In the alternative scenario, proposed by Mabuchi (1971), an across-the-board [ʃ] > [s] change took place except before front vowels.

Although this second hypothesis does not enjoy particularly wide support, it describes a pattern that is not uncommon in Japanese. The resistance of segments to participate in synchronic or diachronic phonological processes is known in the literature as an inalterability effect (Hayes 1986; Inkelas and Cho 1993; Hall 1995). Inalterability effects typically occur with long vowels, geminates, and half-geminates. The phonological descriptions of all of these cases concern features that are shared over the involved segments. The shared features translate into articulatory phonetics as prolonged articulation. The greater articulatory strength makes these forms more resistant to articulatory weakening processes compared to single segments. A well-known example for inalterability effects in Japanese is the case of geminate *–pp–*. A lenition process that changed bilabial stops into fricatives (i.e., [p] > [ɸ]) in Old or Late-Old Japanese failed to apply to long consonants (Frellesvig 2010: 165; Takayama, this volume). Inalterability can also be applied to Mabuchi's (1971) theory about the development of [s]-[ʃ] allophony in Late-Old Japanese. The syllables of [ʃi] and [ʃe] can be analyzed using place features that are shared between consonants and the vowels. This configuration is similar to half-geminates, that is, homorganic nasal-stop clusters, in that only some of the features are shared.

Since both inalterability effects and assimilations are articulatory in nature, it can be claimed that the birth of the sibilant allophones goes back to articulatory motivations no matter which of the two possible changes took place. As (12) illustrate, both types of changes can result in similar allophone distributions.

(12) Two logical ways for developing allophony

(a) Context dependent assimilation

	X		Y
Env$_a$	●		
Env$_b$	○	→	●
Env$_c$	●		

(b) Context free change with inalterability effects

	X		Y
Env$_a$	●	←	○
Env$_b$			●
Env$_c$	●	←	○

4.1.2 Emergence of the /s/ ↔ /ʃ/ contrast

The historical process of creating new phonemes in Japanese often follows a pattern that is referred to by historical linguists as *split* or *secondary split* (Hoenigswald 1960: 93; Antilla 1972: 70). In phonemic splits, allophones that originally were in complementary distribution evolved into phonologically opposing phonemes through changes in their distributions. The contrasts are typically brought about by the elimination of the environment that had been responsible for the allophonic variants in the first place. This type of process is also known as *phonologization* (Campbell 1998: 24). The textbook example for secondary splits is the emergence of front rounded vowel phonemes in Old High German (Anttila 1972: 61; Bynon 1977: 26–27; Kiparsky 1968).

The history of the Japanese sibilant contrast is closely related to the history of Chinese loanwords known as Sino-Japanese words (see Nasu, this volume, and Ito and Mester, this volume). Japanese had established close diplomatic and cultural ties with Chinese by the Heian period (794–1185). Beyond religious and administrative matters, the Japanese writing system and also the sound system reflected this influence. The increased number of loanwords at the time is held responsible for triggering several phonological changes including the emergence of complex onsets (Frellesvig 2010: 184).

The emergence of the /s/ ↔ /ʃ/ contrast was not a single step. Chinese had more consonants than the Japanese consonant inventory could accommodate. Since some of the missing sounds were present as allophones in Japanese, a special adaption method was devised in which foreign syllables were borrowed as a pair of syllables. The first syllable preserved the consonant, while the second one kept the vowel of the original sequence. For example, the Chinese morpheme 'Buddha' 釈 (Modern Japanese [ʃaku]) was adapted as /ʃi/+/a/+/ku/ 志阿久 (Mabuchi 1971: 94). Later the extra onsetless syllable was replaced by a glide-vowel sequence (i.e., /ʃia/ > /ʃija/), which is the phonemic reflection of how vowel hiatuses are resolved phonetically in such environments, even in Modern Japanese (e.g., *piano~pijano*) (see Kubozono, Ch. 8, this volume). The phonologization of the [s]-[ʃ] allophony took place when the syllable pairs simplified into single syllables in Early-Middle Japanese through the reduction of the first vowels: [kija] > [kja], [mija] > [mja], [ɲija] > [ɲa],

[ʃija] > [ʃa]. Similar types of reductions can be observed in Modern Japanese in casual speech: e.g., /okiagaru/ → [okjagaru] ‘to wake up’, /hiotoko/ → [çottoko] (a legendary character), /naɲio/ → [naɲoː] ‘What?’ (Nishida 2001; Kawai 2003). The syllables with the extra glide are referred to by Japanese linguists as “yōon” (拗音), or palatalized syllables, whereas non-palatalized syllables are referred to as “chokuon” (直音). The orthography still represents the palatalized syllables as combinations of two *kana* characters, the second of which is a subscripted kana denoting the glide, as shown in (13).

(13) Japanese palatalized sibilants

シャ	=	シ	+	ヤ
/ʃa/		/ʃi/		/ja/
シュ	=	シ	+	ユ
/ʃu/		/ʃi/		/ju/
ショ	=	シ	+	ヨ
/ʃo/		/ʃi/		/jo/

The opposition within the sibilants first appeared extensively before /a/ (Frellesvig 2010: 199). Most of the oppositions in the context of /o/ and /u/ are results of internal changes in Late Middle Japanese in which vowel coalescence created long /uː/ and /oː/ vowels in massive proportions (e.g., [ʃau] > [ʃoː], [ʃiu] > [ʃuː]) (see Kubozono, Ch. 5, this volume, for more details). By the end of Late Middle Japanese, the /s/ ↔ /ʃ/ contrast was present before the three non-front vowels, leaving two gaps, *[si] and *[se], in the system. At the beginning of Modern Japanese (17th century), [ʃe] turned into [se], shifting the gap under the distribution of [ʃ]. The series of changes are summed up in (14).

(14) Historical development of sibilant contrast up to Modern Japanese

Early Middle Jp		> > >		Late Middle Jp		Modern Jp	
s	ʃ	s	ʃ	s	ʃ	s	ʃ
sa		sa	ʃa	sa	ʃa	sa	ʃa
	ʃi		ʃi		ʃi		ʃi
su		su		su	ʃu	su	ʃu
	ʃe		ʃe		ʃe	se	
so		so		so	ʃo	so	ʃo

The two CV restrictions, *[si] and *[ʃe], with the sibilants in the conservative variety of Modern Japanese can be viewed as the network effect of sibilant allophony combined with the adaptation of Chinese loanwords. The lack of the /ʃe/ syllable in Modern Japanese is in some ways misleading, as /ʃe/ was present in the language up until Modern Japanese when the initial sibilant in the syllable changed into an alveolar fricative (Mabuchi 1971: 136).

4.1.3 Limitations of the sibilant fricative contrast

Modern Japanese witnessed the birth of several new syllable types due to the influence of loanwords. In case of sibilants, however, it was only the /se/ ↔ /ʃe/ contrast that could take root. Although it may be argued that for Japanese speakers, [si] is not impossible to pronounce, several facts suggest that /si/ is not part of the sound system. First, [si] sequences in foreign words are still being adapted into Japanese as [ʃi]. Although there is an orthographic doublet available to accommodate the new syllable type (i.e., /su/+/i/ スィ), the majority of speakers just use /ʃi/ instead: e.g., [ho:muʃikku] 'homesick'. Second, the difficulty of pronouncing [si] is a well-documented characteristic of Japanese language speakers (Vance 1987: 21), presenting a recurring problem in second language education. Finally, even if it is possible to pronounce [si], the /si/ ↔ /ʃi/ opposition is highly confusable, as perceptual experiments testify (Lambacher et al. 2001). All of these arguments question the legitimacy of /si/ in Modern Japanese.

In order to understand the motivations behind the reluctance of the Japanese language to accommodate *[si], it is worth looking at the order of vocalic contexts in which the sibilant contrast emerged. The contrast first appeared in the context of /a/, later it extended to /o/ and /u/. In Modern Japanese it did not cause particularly big difficulty to accommodate the /s/ ↔ /ʃ/ contrast in the environment of /e/ (Kawakami 1977: 46; Vance 1987: 21). Only a handful of examples can be found in which /ʃe/ is not adapted faithfully (e.g., 'shepherd (dog)' → *sepaado*). But even in these cases, variants are available with the faithful /ʃe/ (e.g., *sepaado~shepaado*). Interestingly, [ʃe] is also present in the native vocabulary in emotional expressions: e.g., /ʃee/ (exclamation used by a famous cartoon figure) (Ito and Mester 1995: 830). While opinions may diverge about the phonemic status of /si/ as opposed to /ʃi/ in Japanese, the high vowel is unquestionably the last vocalic environment in which the opposition will emerge. The diachronic order of vowel contexts accommodating the /s/ ↔ /ʃ/ contrast is summarized in (15).

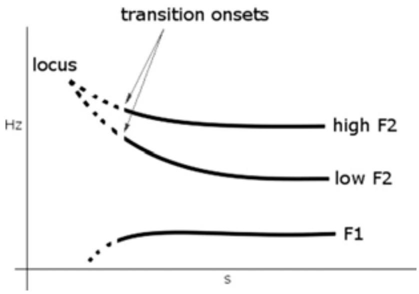
(15) Diachronic order of vowel contexts accommodating the /s/ ↔ /ʃ/ contrast
/a/ > /u/, /o/ > /e/ > /i/

The very fact that the emergence of sibilant contrast did not happen at the same time in all vocalic contexts suggests the influence of vowels on the /s/ ↔ /ʃ/ phonologization process. Phonetic contexts are reported to be an important factor in the perception of phonemic contrasts (Padget 2001; Steriade 2001). Thus, a promising direction for further investigation would be to look into how vowels influence the perception of fricatives.

The two most important acoustic cues that are used to distinguish one fricative from another are the spectral characteristics of the fricatives and formant transitions to neighboring vowels. As for spectral characteristics, Mann and Repp (1980) showed that the noise spectrum of fricative consonants varies with the following vowel. Due to anticipatory lip rounding, fricative noise frequencies are lowered. Since this effect involves both [s] and [ʃ], the perceptual distance between the two consonants is not shrunk, but only shifted. Accordingly, it is hard to relate vocalic effects on spectrum to the diachronic order.

As for formant transitions, the link between vowel color and the extent of F2 transition is obvious. In the case of the post-alveolar or palatal fricatives, the locus of the formant transition is low enough to form a continuous shift leading into the second formant of the following vowel. The locus of the second formant of these sibilants is above 2000Hz, which is close to the second formant of a high front vowel. Since second formants correspond to the frontness of vowels, it can be generalized that the more front the vowel is, the smaller formant transition it maintains. The relation between the level of F2 and the degree of formant transition is depicted schematically in (16).

(16) Schematic F2 transitions in sibilant-vowel sequences



In post-alveolar and palatal fricatives, the formant transitions start within the consonant and continuously lead into the neighboring vowel. In the case of the alveolar sibilant [s], the transition is abrupt, and the second formant does not connect smoothly into the vowel. The differences in formant structure means that [s] and [ʃ] can be distinguished from each other by the presence or absence of the

formant transitions. Since the extent of formant transition is smaller with front vowels, transitional cues are less helpful in this context. Although there is a correlation between the diminishing effect of front vowels and the relatively late emergence of /se/ ↔ /ʃe/ and /si/ ↔ /ʃi/ contrasts in Japanese, it has to be shown that listeners are actually sensitive to transitional cues, as correlation is not a guarantee for casual relationship. It is possible that Japanese speakers ignore transitional cues, and the observed correlation between diachronic and acoustic dimensions is coincidental.

Harris (1958) argued that English speakers are relatively insensitive to transitional cues in the perception of sibilant fricatives. On the other hand, Whalen (1981) showed that formant transitions do contribute to perceptual discrimination. Similar results were reported by Nowak (2006) with Polish speakers. He found that spectral cues are sufficient to identify the tri-partite system of Polish sibilants, although transitional cues can override spectral ones. In the case of Japanese, sibilants are reported to be clearly separable acoustically based on the combination of transitional and spectral cues (Fujisaki and Kunisaki 1978; Funatsu and Kiritani 2000; Hirai et al. 2005). In an experiment with synthetic stimuli, Hirai et al. (2005) demonstrated that Japanese speakers rely more on spectral cues than on formant transitions. This result is supported by experiments with isolated and word-final sibilants in which fricatives are reliably identified – even in the absence of formant transitions (Takeyasu 2009; Matsui 2012: 61–62).

Given these results, however, it is hard see why Japanese university students failed to correctly identify the /si/ ↔ /ʃi/ contrast in Lambacher et al.'s (2001) forced-choice perceptual labeling experiment that used natural utterances. Since in the experiment of Hirai and her colleagues (Hirai et al. 2005) the stimuli consisted of only single, synthesized syllables – a continuum between [sa] and [ʃa] – and the responses were elicited using the AXB method, it is possible that the participants abstracted away from the linguistic content and based their decisions on pure phonetic similarity of the noise portions in the stimuli. In order to directly investigate the influence of transitional cues, a perceptual experiment was a carried out by Pintér (2007) using synthetic CV stimuli. The consonants covered a continuum between [ʃ] and [s], while the subsequent vowel varied on a continuum between [a] and [i]. The results showed that as the vowel height rises, the confusion rate in the identification of [s] vs. [ʃ] also increases, gradually.

Faced with these seemingly inconsistent findings, it is difficult to determine the perceptual strategies of Japanese speakers. Yet, it seems plausible to assume that both spectral and transitional cues can be utilized in the perception of sibilants. Under special circumstances, Japanese listeners can rely solely on spectral cues, but in the presence of vowels the transitional cues outweigh spectral ones (Whalen 1981). Similar perceptual strategies in which transitional cues outweigh consonantal ones can readily explain the relative perceptual difficulties associated with such

oppositions as [ɲi] ↔ [ni], [tsi] ↔ [tʃi], and [hi] ↔ [çi]. A similar effect of transitional cues can be observed in the context of [u], as we will see in the next section.

4.2 Non-sibilant fricatives

The history of non-sibilant fricatives has many features in common with that of the sibilant ones. The emergence of allophones can be attributed to articulatory causes; the phonologization of allophones is greatly hindered in high vowel contexts, presumably because of the poor perceptual licensing properties of these vocalic contexts.

4.2.1 The birth of non-sibilant fricative allophones

It is a widely accepted fact that before Old Japanese, i.e., the 8th century, there were no fricatives other than sibilants in the consonant inventory. The first non-sibilant fricative to appear in the language was the voiceless bilabial fricative [ɸ]. The historical change that produced it is a typologically common lenition process that weakens stops into fricatives. Although the exact time of the change is unknown, it was presumably completed by Early Middle Japanese, or perhaps even earlier (Martin 1987; Frellesvig 2010: 37).

After the [p] > [ɸ] lenition, the development of bilabial fricatives followed different paths in word-initial and intervocalic positions (Hamano 2000; Unger 2004; Takayama, this volume). Intervocalic non-sibilant fricatives turned out to be short-lived, as they went further on the lenition-scale and became sonorants around the 11th century ([ɸ] > [β] > [w]). By the end of Late-Old Japanese, bilabial fricatives could only stand in word-initial position. Traces of this distributional asymmetry can be found in contemporary Japanese. While there are numerous words beginning with /h/, only Sino-Japanese and modern loanwords allow /h/ in intervocalic position tautomorphemically. The number of exceptions to this general rule is minimal (e.g., *haha* 'mother', *ahiru* 'duck', *yahari* 'indeed', *yohodo* 'very') (McCawley 1968: 77–78).

Christian writings (e.g., Rodriguez 1955 [1604]) suggest that the labial fricatives in morpheme-initial position existed at least until the beginning of the Edo period (17th–19th centuries). One of the notable sound changes of Early Modern Japanese was the eventual weakening of voiceless bilabial fricatives to glottal [h]. Bilabial fricatives survived only as a single allophone of /h/ before the high non-front vowel /u/ ([ɸu] *> [hu]). The lenition process was not complete in the context of /i/, either, as fricatives followed by [i] got trapped in the palatal region ([ɸi] > [çi]* > [hi]). The contextually uneven change resulted in an allophonic distribution of non-sibilant fricatives, as shown in (17).

(17) Historical development of sibilant contrasts until Modern Japanese

Late Middle Jp	Modern Jp		
ɸ	ɸ	ç	h
ɸa			ha
ɸi		çi	
ɸu	ɸu		
ɸe			he
ɸo			ho

Both the [ɸ] > [h] lenition, and the resistance to this change can be attributed to articulatory motivations. The [ɸ] > [h] lenition is a common articulatory weakening process in which the pronunciation of the fricative is simplified by losing its oral place of articulation. This and similar weakening processes are responsible for the birth of [h] in several languages (Lass 1984: 179; Foulkes 1997). The reluctance of [ɸu] and [çi] to weaken into [hu] and [hi] respectively is similar to the blocking of [ʃ] > [s] in the context of front vowels. In all of these cases, the consonant and the vowel have shared place features. The shared labiality[9] in the case of [ɸu] and the shared palatal feature in the case of [çi] are responsible for the blocking of weakening.

A detail that was omitted from the above explanation is the [ɸ] > [ç] change before the high front vowel. While this change may suggest the presence of an interim palatal stage in the weakening process [ɸ] > [ç] > [h], it is more likely that the lenition of the bilabial fricative involved a loss of its place specification (or "de-oralization" in Lass' (1984: 179) terminology). From both phonetic and phonological points of view, it is reasonable to treat [h] as a segment that lacks place specification on its own (Laver 1994: 245; Ladefoged and Maddieson 1996: 322–326). Most likely, the change took place in a gradual fashion, in which the disappearing labiality gradually yielded to the palatal feature from the neighboring high vowel. It was not only the high front vowel that spread its place feature to the weakening bilabial fricative. Palatalized bilabial fricatives (i.e., [ɸj]) were another source of palatal fricatives, as we will see in the next section.

4.2.2 Emergence of the / h / ↔ / ç / contrast

Under the influence of Chinese loanwords, bilabial fricatives, just like sibilants, developed palatalized variants, although to a lesser extent. The syllables of /ça/, /çu/,

9 Although Japanese /u/ is not rounded, it is not completely unrounded, either.

and /ço/ in Modern Japanese with palatal fricatives originated from the complex [ɸjV] sequences, which in turn arose from the earlier [pjV] forms. These early [pjV] forms emerged under the influence of Chinese loans in Early-Middle Japanese (Frellesvig 2010: 184, 314). As Christian writings testify, the palatalized fricative pronunciation, e.g., [ɸjaku] 'hundred', existed at least until the beginning of Modern Japanese (17th century). Since the merger of the complex onset consonants [ɸj] removed the palatal glide from the output, the change created a direct opposition between /h/ and /ç/ (e.g., [ɸjaku] > [çaku] 'hundred' versus [ɸaku] > [haku] 'white'). The contrast did not emerge in those contexts where the contrast of palatalized versus non-palatalized onsets was absent in the first place (i.e., /Ci/ = /Cji/, /Ce/ = /Cje/). Before the high front vowel, the palatalization process resulted in a non-contrastive [çi] syllable. Before the front mid vowel, the bilabial fricative underwent lenition, i.e., [ɸe] > [he]. The absence of [çe] forms was a natural consequence of the historical absence of /Cje/ forms. As shown in (18), /ç/ participates in fricative contrasts only in the contexts of /a/, /u/ and /o/.

(18) The emergence of the /h/ ↔ /ç/ contrast

Modern Jp		
ɸ	ç	h
	ça	ha
	çi	
ɸu	çu	
		he
	ço	ho

Lexical borrowings after the Meiji Restoration did not significantly affect the distribution of the palatal fricative, as this phoneme was not dominant in the donor languages. The palatal fricative [ç] is not a phoneme in English, but it does surface in the adaptation of /hj/ sequences, such as in 'huge' → *hyuuji* [çu:dʒi], 'human' → *hyuuman* [çu:man] or 'humor' → *hyuumoa* [çu:moa]~*yuumoa* [ju:moa]. In German, the palatal affricate has allophonic status: its distribution is restricted to coda position in complementary distribution with [x] (Wiese 1996). As shown in (19), German word-final palatal fricatives are often borrowed into Japanese as singleton or geminated [ç] followed by [i] (Tews 2008).

(19) Loanwords with palatal fricatives in contemporary Japanese

(Marlene) Dietrich	*diitorihhi*	[diːtoriççi]
(Paul) Tillich	*tirihhi*	[tiriççi]
Ludwig (van Beethoven)	*rudobihhi*	[ruːdobiççi]
Pfennig	*penihhi*	[peɲiççi]
Zürich	*chuurihi~chuurihhi*	[tʃuːriçi]~[tʃuːriççi]

German provides another phonologically interesting source for palatal fricatives. Sequences of /h/ plus front rounded high vowels in German are often adapted into Japanese as [çu] (e.g., 'Hütte'$_{\text{GERMAN}}$ → *hyutte* [çutte] 'mountain hut'). Although there are only a handful of loanwords showing this process, they highlight an intriguing combination of phonological unpacking and merger.[10] The [high] feature from the front rounded vowel is unpacked and merged with the preceding consonant, resulting in a palatal fricative (Dohlus 2004).

4.2.3 Emergence of the /h/ ↔ /ɸ/ contrast

The third phoneme in the non-sibilant fricative system is a recent development. The bilabial fricative /ɸ/ emerged in response to the increasing pressure from loanwords that proliferated in the language at a huge rate after the Meiji Restoration in 1868. Owing to its closeness in time, the phonologization of [ɸ] is exceptionally well-documented. Pre-war studies report that the proper pronunciation of [ɸ] was not particularly common; it was a hallmark of good education. People who had difficulties with this sound resolved [ɸV] forms either as a sequence of the conservative [ɸu] plus vowel sequence, e.g., 'film' → [ɸu.i.ru.mu] (Arakawa 1932: 218–229), or replaced the bilabial fricative with the native /h/ phoneme, e.g., 'koffie'$_{\text{DUTCH}}$ → *koohii*, 'filet'$_{\text{FRENCH}}$ → *hire*, 'platform' → *purattohoomu* (Umegaki 1944: 141, 210). These resolution strategies are mostly obsolete now. The use of /ɸ/ is well integrated into contemporary Japanese, as illustrated in (20).

(20) Loanwords with [ɸ] in contemporary Japanese

/ɸa/	*fakkusu*	'fax'	*fairu*	'file',
/ɸi/	*firumu*	'film'	*gurafikku*	'graphic'
/ɸu/	*fukku*	'hook'	*fuudo*	'hood' or 'food'
/ɸe/	*kafee*	'cafe',	*feaa*	'fair'
/ɸo/	*sumaatofon*	'smart phone'	*foomaru*	'formal'

10 The term 'unpacking' refers to a phenomenon in loanword phonology in which a foreign sound that is absent from the borrowing language is adapted as a sequence of existing phonemes (e.g., front rounded high vowel [y] → /ju/) (Paradis and Prunet 2000; Kubozono, Ch. 8, this volume).

4.2.4 Limitations of the non-sibilant fricative contrasts

Comparing the conservative and the innovative varieties of Japanese (21), it can be seen that there are still many phonotactic gaps in the non-sibilant fricative system. Neither /h/ ↔ /ɸ/ nor /h/ ↔ /ç/ is completely phonologized even today. The /h/ ↔ /ɸ/ opposition is still absent before /u/, while the /h/ ↔ /ç/ opposition has not managed to extend to the contexts of /i/ and /e/. Neither the writing system nor the spoken language show any signs of accommodating these absent contrasts.

(21) Conservative and innovative syllables with non-sibilant fricatives

Conservative				Innovative		
ɸ	ç	h		ɸ	ç	h
	ça	ha		ɸa	ça	ha
	çi			ɸi	çi	
ɸu	çu			ɸu	çu	
		he		ɸe		he
	ço	ho		ɸo	ço	ho

The pattern expressed by the gaps has several features in common with sibilant fricatives. The lack of /h/ ↔ /ç/ contrast before front vowels is in parallel with the lack of sibilant /s/ ↔ /ʃ/ opposition in the same environment in the conservative variety. Similar to sibilants, the neutralization of the palatal fricative and [h] before high vowels can also be attributed to the effect of diminished salience of transitional cues.

The palatal fricative [ç], although weaker in energy, has similar spectral characteristics to Japanese [ʃ] (Yamazaki, Tsugi, and Pan 2004). The bilabial fricative, unlike the sibilant and palatal ones, has a flat spectrum, and it is apparently weaker. All of these fricatives are associated with clear formant structure in that formant transitions into adjacent vowels are salient acoustic and perceptual features for all of them. In sharp contrast with these fricatives, the glottal [h] sound is often characterized as a whispery sound, a breathy onset to a vowel (Koizumi 1996; Jōo 1982; Gordon et al. 2002). Even its categorization as a fricative is questionable (Laver 1994: 245; Ladefoged and Maddieson 1996: 322–326; Crystal 1980: 24).

Accordingly, the oppositions the glottal [h] is engaged in with the palatal and the bilabial fricative can rely on divergent formant trajectories and on differences in spectral characteristics. Following the arguments laid down in the discussion of sibilants (section 4.1.3), the high-vowel contexts are assumed to be disadvantageous also for the [h] versus fricative contrasts because of the diminished salience of

formant transitional cues. The front vowels are dis advantageous for the palatal-glottal opposition, because these vowels maintain relatively small transitions in combination with the palatal fricative. The high non-front vowel shares a labiality with the bilabial fricative, which results in static-like formant trajectories. In both of these cases, the formant transitions are not distinct enough to tell them apart from the non-altering trajectories of [hV] sequences. In the absence of transitional cues, the listeners have to resort to less reliable spectral cues.

Although the experiments with sibilants above prove that Japanese speakers are capable of relying on spectral characteristics, they do not seem to utilize this perceptual cue fully with sibilants. The same logic applies to non-sibilant fricatives. In addition, the non-sibilant fricatives are acoustically weaker than sibilant ones, so their spectral differences provide even less support for perceptual distinction. Based on the strength of fricatives, it can be predicted that the /si/ ↔ /ʃi/ contrast is likely to emerge before the /hi/ ↔ /çi/ or /hu/ ↔ /ɸu/ contrasts if future generation of Japanese speakers accommodate spectral-cue-oriented perceptual strategies.

The assumption that Japanese speakers rely more on formant transitional cues predicts perceptual difficulties also with /ʃi/ and /çi/. The points of articulation for alveo-palatal [ʃ] and palatal [ç] are quite close. When followed by [i], both [ʃi] and [çi] maintain a relatively flat formant trajectory at around the same frequency level, which can cause perceptual confusion. This prediction seems to be borne out in less prestigious varieties of Japanese. Vance (1987: 22) reports a merger of /ç/ and /ʃ/ syllables in some nonstandard dialects. Interestingly, most of the examples cited, which have been replicated in (22), describe merger before the vowel /i/.

(22) The ongoing [ʃ]-[ç] merger in nonstandard dialects

hibachi	[ʃibatʃi]~[çibatʃi]	'brazier'
hidoi	[ʃidoi]~[çidoi]	'terrible'
shiku	[ʃiku]~[çiku]	'spread out'
shichi	[ʃitʃi]~[çitʃi]	'seven' (added by the author)

The reason why alveo-palatal and palatal fricatives do not merge in most other dialects is that spectral differences between sibilant and non-sibilant fricatives are presumably robust enough to maintain a perceptual demarcation line.

The only gap in the fricative system that does not involve high vowels is due to the lack of the /he/ ↔ /çe/ opposition. Since words with /hje/ sequences were originally absent when the [ɸj] > [ç] change took place, the absence of [çe] syllables can be interpreted as a natural, accidental gap. For similar reasons, /ʃe/ was also absent at the beginning of Modern Japanese. However, while the gap of /ʃe/ was filled in relatively easily, there are no [çe] syllables in the language even today. This is due partly to the low frequency of [çe] sequences in the donor languages. It is also due to perceptual difficulties predicted by similar formant trajectories between [çe] and [he]. The last example in (23) demonstrates a case where input [tçe] is

adapted as geminate [tʃe], merging stop and palatal features from the input: i.e., 'Mädchen'$_{\text{GERMAN}}$ [metçen] → *metchen* [mettʃen].

(23) Loanwords from German with [çe] adapted as *he* and *che*

'Märchen'	*maruhen*	'fairy tale'
'München'	*myunhen*	*German city*
'Mädchen'	*metchen*	'girl'

The fact that the emergence of the /se/ ↔ /ʃe/ opposition precedes that of the /he/ ↔ /çe/ contrast is in accordance with the assumption that differences in spectral characteristics are more salient in sibilants than in non-sibilants.

4.3 The stop/affricate system

The stop/affricate system refers to the tripartite contrast of alveolar stop [t], affricate [ts] and [tʃ] in Modern Japanese. The development of the stop/affricate system shows both similarities and differences to that of fricatives.

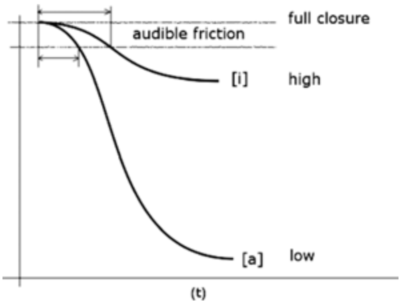
4.3.1 Early development of the stop/affricate system

The discussion of alveolar stops and affricates goes back to Late Old Japanese. Opinions converge in assuming a single phoneme as the ancestor of contemporary Japanese [t], [ts] and [tʃ] (Kindaichi 1932: 190; Martin 1987; Frellesvig 2010: 37). There are several internal and external pieces of evidence that suggest that /t/ and /d/ were uniformly pronounced as [t] and [d] in all vocalic contexts until the end of Middle Japanese. For example, a late 15th century Japanese textbook written in Korean (Hamada 1952) describes the syllables corresponding to contemporary Japanese [tʃi] and [tsu] without affrication (Hamada 1952: 26–27). Also, in some dialects of Shiga and Toyama Prefectures, [ti] and [di] remained un-assibilated (Martin 1987: 16).

The development of affricate allophones was greatly influenced by Chinese loanwords. Similar to the fricative system, Chinese loanwords introduced palatalized variants of alveolar stops mainly before the vowel [a] (Frellesvig 2010: 200). It is difficult to reconstruct the exact phonetic values of the palatalized forms but presumably they were pronounced as [tija] or [tja]. The appearance of affricates in Japanese was due to an affrication process that affected palatalized syllables and alveolar stops before high front vowels (i.e., [tj] > [tʃ] , [ti] > [tʃi], and [tu] > [tsu]). The time of the assibilation of [ti] and [tu] is uncertain, but it is assumed to be sometime around the Muromachi period (14th–16th centuries) (Hashimoto 1950: 88; Martin 1987: 16). The motivation for the change is apparently articulatory.

During the release phase of a stop, the air built up behind the closure escapes rapidly, within a few milliseconds. If the release is slower, the articulators spend more time in close proximity to each other, which prolongs the duration of the burst. A longer burst produces audible friction, which is perceptually identified as a fricative. Thus, insufficient coordination by the speaker results in a stop followed by a fricative, that is, an affricate. The trajectories for tongue movements in [ti] and [ta] are represented schematically in (24).

(24) Schematic trajectories of tongue movements in [ti] and [ta]



This type of assibilation is more likely to occur before high vowels or palatal glides for articulatory reasons (Jaeger 1978: 316). In a transition from a stop to a high vowel, the tongue spends more time in the zone that causes audible friction. If the transition targets a low vowel, the tongue passes through the friction zone relatively rapidly. This articulatory difference predicts that under similar conditions stops have longer friction components before high vowels than before lower ones. The common diachronic changes of [ti] > [tʃi], [tj] > [tʃ] and [tu] > [ʦu] reflect the extra friction noise that high vowels produce.

The assibilation in Japanese did not create an even distribution of affricates. For obvious reasons, affricates appeared in the greatest numbers before the high vowels (i.e., [tʃi] and [tsu]). Since palatalized consonants or “yōon” were present mainly in the context of /a/, the next most frequent affricate syllable was [tʃa]. Most syllables with [tʃoː] and [tʃuː] in Modern Japanese, as the long vowel may imply, are results of vowel coalescence which took place around Late Middle Japanese (e.g., 町 [tʃau] > [tʃoː] ‘town’; 宙 [tʃiu] > [tʃuː] ‘space’) (Kubozono, Ch. 5, this volume, for more details). The difficulty to find Sino-Japanese words with short /o/ and /u/ vowels following affricate [tʃ] is related to the scarcity of palatalized [tjo] and [tju] syllables in Middle Japanese. Likewise, the complete lack of [tʃe] syllable goes back to the absence of [tje] in contemporary Chinese loanwords.

4.3.2 Emergence of contrast in the stop/affricate system

The internal assimilatory changes resulted in three pairs of oppositions in the stop/affricate system by the beginning of Modern Japanese: /tʃa/ ↔ /ta/, /tʃu/ ↔ /tsu/, and /tʃo/ ↔ /to/ (Frellesvig 2010: 385). The resulted system is outlined in (25).

(25) Stop/affricate system in the conservative dialect (surface forms)

Conservative		
t	tʃ	ts
ta	tʃa	
	tʃi	
	tʃu	tsu
te		
to	tʃo	

Although assimilatory changes did not produce [tʃe] syllables, their absence from Early Modern Japanese was not systematic. The accidental nature of the absence of [tʃe] is supported by the speed and the ease with which loanwords filled this gap. No perceptual or articulatory difficulties are reported with innovative /tʃe/ forms. The new syllable occurred in such loanwords as ‘Nietzsche’$_{\text{GERMAN}}$ → *niiche*, ‘check’ → *chekku*, or ‘cello’$_{\text{ITALIAN}}$ → *chero*. Interestingly, [tʃe] is not restricted to loanwords, but is present in the natively-coined exclamative word *che* ‘damn’ (Ito and Mester 1995: 830). The ease of adapting /tʃe/ is comparable to the ease of integrating /ʃe/ into the language.

The order in which syllable types after *che* (e.g., *ti, tu, tsa*… etc.) appeared in Japanese is subject to debate. Some of the changes are still in progress. One source of data that can help outline a rough diachronic order is presented to us by reforms of the writing system. Although it is generally ill-advised in a phonological study to rely on spelling practices, these reforms provide valuable, albeit not unconditionally reliable, insight into the native speaker’s intuition about the acceptability of novel syllable types.

The Japanese National Language Committee (*Kokugo Shingikai*) prepared a list of proposals in 1954 regarding the kana transliteration of foreign words in an attempt to unify and standardize the chaotic transcription practices (Ishiwata 2001). In general, the proposals encouraged the use of conservative spellings, but some new combinations of characters were also acknowledged for established loanwords. Two new syllables that were acknowledged in the stop/affricate system were /tʃe/ (チェ) and /ti/ (ティ). In contrast with /tʃe/, the adaptation of /ti/ sequences in

Japanese was not without hesitation. As shown in (26), [ti] used to be systematically avoided in loanwords. This instability of /ti/ is fossilized in older loanwords in which foreign [ti] sequences are borrowed either as /tʃi/ (チ) and, less frequently, as /te/ (テ) (see Kubozono, Ch.8, this volume).[11] Since it is almost impossible to find loanwords in which /tʃe/ is adapted in a conservative manner, it is reasonable to say that /tʃe/ became an acknowledged syllable before /ti/.

(26) Examples for the adaption of [ti] as *chi* and *te*

'team'	→	*chiimu*	'sticker'	→	*sutekkaa*
'ticket	→	*chiketto*	'destination'	→	*desuteneeshon*
'romantic'	→	*romanchikku*	'Tim'	→	*chimu, temu*
'tilde'	→	*chiruda*	'CVT'	→	*shiibuitee*[12]

The increasing discrepancy between actual usage and the principles laid down in 1954 necessitated a revision of the official spelling recommendations. In 1991 the Japanese National Language Committee published another proposal on the transcription of foreign words. In this new proposal, there is a clear distinction between established syllables and hesitating ones. For example, /ti/, /tsa/, /tse/, and /tso/ (ティ, ツァ, ツェ, ツォ) are regarded as accepted syllable types in Standard Japanese, whereas /tu/ (トゥ) and /tsi/ (ツィ) remain in the gray zone. The proposal advises explicitly against the use of /tu/ and /tsi/ in the written language, and suggests replacing them with conservative /tsu/ (ツ) and /tʃi/ (チ) syllables, respectively. The changes between conservative and innovative dialects are outlined in (27).

(27) Development of alveolar stop and affricates in Modern Japanese

Conservative			> >			> >			Innovative		
t	tʃ	ts	t	tʃ	ts	t	tʃ	ts	t	tʃ	ts
ta	tʃa		ta	tʃa		ta	tʃa	tsa	ta	tʃa	tsa
	tʃi		ti	tʃi		ti	tʃi		ti	tʃi	tsi
	tʃu	tsu		tʃu	tsu		tʃu	tsu	tu	tʃu	tsu
te			te	tʃe		te	tʃe	tse	te	tʃe	tse
to	tʃo		to	tʃo		to	tʃo	tso	to	tʃo	tso

11 There are several examples that avoided [di] by replacing the vowel with [e:]: '(the letter) D' → *dee*, 'candy' → *kyandee*, 'CD' → *shiidee*.

12 From TV commercial *Teinenpi Shōjo Haiji*, Part 4. 0:50. CVT stands for Continuously Variable Transmission.

Acknowledging various /tsV/ syllables in the language hallmarks the emergence of a new affricate phoneme /ts/. Since /ts/ is absent as a phoneme in English, which is the main donor language, the number of words containing [ts], other than [tsu], is relatively low. The main sources of the voiceless alveolar affricates are German, Italian, and Slavic languages (e.g., 'Zeitgeist'$_{\text{GERMAN}}$ → *tsaitogaisuto*, 'canzone'$_{\text{ITALIAN}}$ → *kantsoone*, 'tzar'$_{\text{RUSSIAN}}$ → *tsaaru*). Despite their rarity, most /tsV/ sequences are perceptually and articulatorily stable. Although there are some variants of *tsa* such as *tsu.a* (ツア) or *za* (ザ) (e.g., 'Mozart' → *mootsaruto~mootsuaruto*, 'pizza' → *piza*), these forms are uncommon. Furthermore, the syllable *tsa* and *tso* can also occur even in native words. The intimate variant of 'father' *otoosan* is *ottsan*; the non-standard variant of *gochisoosama* 'feast' is *gottsan* (Vance 1987: 23).

The last two syllables to appear in the stop/affricate series are *tsi* and *tu*. As for *tu*, recent studies tend to accept it as an established innovation in Japanese, although this view is not unanimous. The instability of *tu* is vividly reflected in the vocabulary, as most of the loanwords with original [tu] sequences are adapted with conservative [tsu] (e.g., 'two' → /tsuu/, 'tour' → /tsuaa/). In some cases an innovative variant of [tu] is also available (e.g., 'Bantu' → /bantuu/). Even though the writing system encourages the use of the innovative variant in newer loanwords, and also many speakers can pronounce [tu], there is always an option to fall back to the more comfortable conservative [tsu] variant. This level of hesitation between conservative and innovative forms is not observed with /tʃe/ or /ti/.

The last syllable in our discussion is /tsi/. While words with *tu* are not difficult to find, loanwords with *tsi* are extremely rare partly because of its rarity in the donor languages in the first place. The examples cited in the literature such as *mittsi* 'Mitzi Gaynor' (Vance 1987: 23) and *eritsin* 'Yeltsin' (Ito and Mester 1995: 826) have marginal presence. Additionally, these words almost always have a variant without *tsi* (e.g., 'paparazzi' *paparattsi~paparacchi*, 'Zyklus'$_{\text{GERMAN}}$ → *tsikurusu~chikurusu* 'cycle').

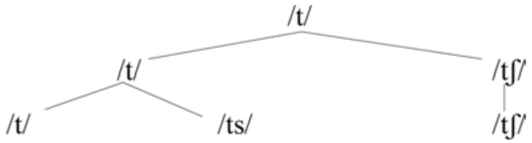
4.3.3 Limitations of the stop/affricate contrasts

The first opposition in the voiceless non-fricative obstruent series emerged through the bifurcation of /t/ and /tʃ/ phonemes. Because of the particular distribution of palatalized syllables in Middle Japanese, the phonologization of affricate /tʃ/ initially happened before the low vowel, and then extended to the context of /o/ and /u/ not much later. Western loanwords gave birth to the /te/ ↔ /tʃe/ opposition relatively early in Modern Japanese. Finally, the /ti/ ↔ /tʃi/ contrast emerged. In terms of vocalic contexts, the emergence of the /t/ ↔ /tʃ/ contrast, summarized in (28), shows close resemblance to the development of sibilant fricatives of /s/ and /ʃ/.

(28) The diachronic order of vowel context accommodating the /t/ ↔ /tʃ/ contrast
/a/ > /u/, /o/ > /e/ > /i/

After the binary opposition of /t/ and /tʃ/ appeared in all vocalic environments, the alveolar non-fricative obstruent series underwent another split creating a tripartite opposition of /t/, /tʃ/ and /ts/, as shown in (29).

(29) The historical development of voiceless alveolar stop and affricates



Determining the order of vocalic context for the three-way distinction is not straightforward because of the rarity of words with *ts* (other than *tsu*). The number of words with *tu* in Japanese dictionaries is higher than the number of entries with *tsi*; still *tu* is considered to be less stable. Besides the arguments presented above related to writing reforms and the variations in loanwords, there is another case that suggests the phonotactic instability of *tu*. In the name of a Serbian politician, 'Koštunica', both [tu] and [tsa] are present but only [tsa] is borrowed faithfully, while [tu] is replaced by [to]: 'Koštunica' → *koshutonitsa*. It is only *tsi* among the tsV syllables whose phonotactic instability is comparable to *tu*. Based on these data it is difficult to go into further details, but it can be claimed that the three-way opposition in the stop/affricate system prefers the context of non-high vowels over high ones, i.e., /a/, /e/, /o/ > /u/ /i/.

A brief overview of perceptual factors explains why the two-way contrast was hindered before front vowels and why the three-way opposition is sub-optimal in high-vowel contexts. There are three major acoustic cues responsible for maintaining the difference between /t/, /tʃ/ and /ts/. First, the duration of the friction noise following stop closures is instrumental in cueing the stop-affricate contrast. The plain stop, i.e., /t/, is followed by a relatively short release burst of around 40ms, whereas affricates have a long noise component (Shaw and Balusu 2010). Second, the spectra of the friction noise following the closure help differentiate between the two types of affricates, i.e., [tʃ] ↔ [ts]. Third, format transitions to consecutive vowels help single out [tʃ], as the fricative component [ʃ] has a clear formant locus (see above, but also Raphael 2005 and Hall et al. 2006). The major acoustic cues in the stop/affricate system are summarized in (30).

(30) Available acoustic cues in stop/affricate contrasts

	t ↔ ts	t ↔ tʃ	tʃ ↔ ts
noise duration	O	O	
noise spectrum	Δ	O	O
formant transition		O	O

A brief comparison of acoustic features over the pairwise combination of consonants shows that the [t] ↔ [tʃ] opposition can utilize the greatest number of cues. It is an interesting coincidence that this is the first opposition to emerge historically in the stop/affricate system. The least salient opposition seems to be maintained between /t/ and /ts/, but both the weighting of cues and the contextual effects on perception complicate the picture.

As it was pointed out in the discussion of alveolar versus palatal fricatives, transitional cues are the weakest in the context of /i/. So /tʃi/-/tsi/ and /ti/-/tʃi/ represent the least favorable oppositions in this respect. Spectral cues show reduced utility in the context of /u/. Spectral differences measured in the release burst of [t] and [ts] are found to be the weakest in the environment of /u/ (Hall et al. 2006). As for the durational cue, the same articulatory model applies as in the explanation of assibilation. The duration of the stop release noise is significantly longer preceding high vowels than preceding other vowels, which makes the stop-fricative border less prominent in the context of /i/ and /u/. The least favorable contrasts are summarized in (31).

(31) Least favorable vocalic context for opposition (if available)

	t ↔ ts	t ↔ tʃ	tʃ ↔ ts
noise duration	*ti-tsi *tu-tsu	*ti-tʃi *tu-tʃu	*ns*
noise spectrum	*tu-tsu		
formant transition	*ns*	*ti-tʃi	*tʃi-tsi

The disadvantage of the /tu/ ↔ /tsu/ contrast is rather apparent as /u/ is listed as the least favorable context for both of the cues the /t/ ↔ /ts/ contrast relies upon. The opposition of /ti/ ↔ /tʃi/ may seem to be similarly disfavored, but in this case spectral cues are also available, making the opposition more salient than /tu/ ↔ /tsu/. It is questionable whether spectral cues are available to cue the /tʃi/ ↔ /tsi/ contrast. In keeping with a pattern similar to that of sibilants (i.e., /ʃi/ ↔ /si/), it is possible that transitional cues override spectral ones.

Presumably, the perceptual analysis raises as many questions as it answers. Nevertheless, it cannot be denied that the phonotactic hierarchy of CV constraints, which is based on phonological behavior of syllables, shows correlations with the perceptual characteristics of the CV sequences.

5 Representations

The question of how consonants should be represented is a long standing issue in Japanese phonology with no clear consensus on the horizon. Should the syllable

[ʃa] be represented as /ʃa/, /sia/, or /sja/, or something else? Does [tsu] go back to /tu/ or /tsu/? Different phonological theories give different answers to these questions. The following points enumerate some of the most common representational practices with special focus on their treatment of emerging consonants and phonotactic gaps. Without questioning the value of phonetic approaches, it is intriguing to see how phonology addresses the issue of emerging consonants in Japanese. It is possible that some phonotactic restrictions are directly related to how native speakers store and manipulate speech sounds and how, for that matter, phonological structures are represented in the brain.[13]

5.1 Early non-derivational approaches

Relying on the distinctive function of the Saussurean sign, phonemes can be defined as the minimal pronounceable units that contrast meaning. If a minimal pair or an appropriate near minimal pair can be found for two sounds in the language, then they are considered to be separate phonemes. For example, *sori* 'sled' and *shori* 'treatment' are separate words in Japanese, so /s/ and /ʃ/ are to be treated as distinct phonemes.

Bernard Bloch, a leading figure of the American Descriptivist school (Blevins 2013), published a series of articles about Japanese involving a thorough analysis of the Japanese sound system (Bloch 1950). After identifying the phonetic forms and their environments, Bloch proposed 23 consonant phonemes for the language (Bloch 1950: 113). Being aware of the numerous co-occurrence restrictions between consonants and vowels, he presented the syllabary in (32) as a summary of underlying forms (Bloch 1950: 119).

(32) Underlying CV units in Modern Japanese (Bloch 1950: 119)

	/s/	/ʃ/	/h/	/ç/	/t/	/tʃ/	/ts/
a	sa	ʃa	ha	ça	ta	tʃa	
i		ʃi	hi	çi		tʃi	
u	su	ʃu	hu	çu		tʃu	tsu
e	se		he		te	tʃe	
o	so	ʃo	ho	ço	to	tʃo	

13 Note that representations in generative frameworks are highly abstract entities. Derivations are not believed to correspond to any articulatory, perceptual, or cognitive processes.

Bloch's analysis has at least two important theoretical implications. First, it assumes that consonant-vowel restrictions are enforced at the underlying level. For example, the vowel phoneme /i/ cannot follow the phoneme /s/ in the underlying representation: */si/. Second, closely related to the previous point, the underlying forms are faithful to the surface representation even in those cases where other analyses would automatically treat surface forms as positional allophones. For instance, in Bloch's analysis [ʃi] is the phonetic realization of /ʃi/, not the positional allophone of /si/. In contrast, the voiceless bilabial fricative [ɸ] is not treated as a phoneme because it is seen as being in free variation with [h] in the context of /u/ (i.e., /hu/ → [hu]~[ɸu]) (Bloch 1950: 108). The syllable inventory provided by Bloch may seem to be too restrictive, but further phonemes such as the affricate /ts/, the bilabial fricative /ɸ/, and its voiced counterpart /v/ are also acknowledged in the innovative dialect (Bloch 1950: 122).

In this early structuralist, non-derivational approach, the emergence of consonant contrasts is rather straightforward. For example, the emerging /tu/ in contemporary Japanese can be analyzed as the emergence of an underlying /tu/ syllable. While this approach is simple and straightforward, it does not provide any clues as to why certain gaps in the syllabary are more difficult to fill than others.

5.2 Derivational approaches

Probably the most widely used techniques for representing the consonants of Japanese are derivational ones. Derivational frameworks in phonology assume that surface forms are derived from underlying representations by rules or other mechanisms. Although it is possible in derivational frameworks to give an analysis for Japanese consonants that is analogous to that of Bloch's, there are several arguments against doing so. First, the descriptivist analysis can be criticized for positing ad-hoc CV restrictions at the underlying level. For example, if both /t/ and /u/ are full-fledged phonemes, it is reasonable to expect their combination at the beginning of syllables (Tsujimura 2007: 36). Second, surface variations in inflectional morphology are difficult to deal with if underlying representations merely copy surface forms. As shown in (33), the consonant variation in verb paradigms can make the identification of the verb roots problematic.

(33) Underlying CV units in Modern Japanese (Bloch 1950: 119)

	'wait'				'speak'			
negative	ma	t	–	anai	hana	s	–	anai
conjunctive	ma	tʃ	–	imasu	hana	ʃ	–	imasu
non-past	ma	ts	–	u	hana	s	–	u
imperative	ma	t	–	e	hana	s	–	e
presumptive	ma	t	–	oo	hana	s	–	oo

The issues of underlying CV restrictions and the morphophonemic alternations can be elegantly addressed by the introduction of derivational rules. The constraints on the underlying CV sequences can be removed because rules convert the illicit CV forms to well-formed surface structures (e.g., /tu/ → [tsu], /si/ → [ʃi]). This approach has the additional benefit of reducing the complexity of the consonant inventory. Through assimilation rules several consonant phonemes can be eliminated. For example, surface [ʃ] can be analyzed as the output of a /si/ → [ʃ] or /sj/ → [ʃ] assimilation. The morphological analysis also benefits from assimilation rules as it allows for uniform underlying representations for words with root-final consonant variation (e.g., /hanas-/ for [hansanai], [hanaʃi]). Further justification for allophonic rules can be found in historical changes (see [ti] > [tʃi] above), and in loanword phonology (e.g., 'ticket' → [tʃiketto]).

With all its merits, the derivational approach is not particularly suited for modeling changes in the history of consonants. First of all, from a derivational point of view, there are no new consonants in Japanese since Old Japanese. The problem of emerging consonant phonemes does not exist at all – or is a misnomer at best. Relying on assimilation rules, the innovative consonants can be analyzed as linear sequences of conservative segments. The sibilant fricative /ʃ/ can be viewed as /sj/, the affricate [tʃ] as /tj/, the palatal fricative [ç] as /hj/, and so on. With the application of this representational unpacking, the consonant inventory is practically assumed to have stuck in Old Japanese. The innovations concern only the linear arrangement of existing phonemes.[14]

Second, the historical process of phonologization is also problematic. New consonants and syllables are modeled similarly to the structuralist approach, by the introduction of new (arrangement of) phonemes at the underlying level. This analysis becomes complicated if an emerging syllable coincides with the input of an assimilation rule (see *sukima* in section 3.1 above). For instance, in order to accommodate *ti* as a new syllable, the /ti/ → [tʃi] assimilation has to be suspended. But this rule is still needed in the phonology. This contradiction can be solved either by claiming that the assimilation rule is fossilized (Hattori 1960, but see Nishida 2010), or by assuming that the assimilation rule does not apply on the lexical domains of loanwords (McCawley 1968: 62–75). During the fossilization process the surface [tʃi] forms are re-analyzed as /tʃi/. The main problem with this view is that it is difficult to see how the introduction of a new syllable such as *ti* can possibly cause the phonological re-analysis of other well-established syllables. The domain-specific application of rules is also questionable as it gives too much power to the

14 The dangers of unbounded abstraction are realized as a problem in the generative literature. A possible solution is to disallow underlying forms that have no phonetic realization on the surface (Kiparsky 1968; but see Ségéral and Scheer 2001). The linear decomposition of innovative consonants in Japanese fails to fulfill this requirement because the hypothesized sequences, such as [sj] or [tj], never surface.

analysis. Although neither of these solutions is particularly elegant, it can be argued that they are needed exactly for those syllables whose introduction into the language is not smooth. Thus, difficulties at the level of formal description can be interpreted as the expression of cognitive challenges involved in the adaption of new phonotactic patterns.

5.3 Underspecification approaches

Without going into details about intricacies of underspecification theories (see Krämer 2012), underspecified segments here refer to abstract units that encompass more than one phoneme by leaving some distinctive features unspecified (on the related notion of archiphonemes, see Krämer 2012: 16; Anderson 1985: 107; and Akamatsu 2000: 19). For example, the underspecified sibilant /S/ encompasses both members of the otherwise contrastive /s/ ↔ /ʃ/ opposition. The missing features are supplied either by the phonetic environment or by default values. As for /S/, the palatal place of /i/ is borrowed before the high front vowel (i.e., /Si/ → [ʃi]); in other environments the default coronal place gets inserted (e.g., /Sa/ → [sa]).[15]

The use of underspecified segments can elegantly handle variations in inflectional morphology (e.g., os+ 'push': /oS+i/ → [oʃi], /oS+u/ → [osu]) and explain the gaps in the syllabary. The phonologization of allophonic relations can be modeled by introducing new underlying consonants with their place features specified. So, the phonologization of the [t]~[tʃ] allophony can be analyzed as the emergence of a /tʃ/ phoneme supplementing the existing /T/. Similarly, the appearance of [ti] introduces a fully specified /t/ phoneme besides underlying /tʃi/ and /T/. Using three underlying segments may seem to be nonsensical, especially because the surface [tʃi] can be analyzed either as /Ti/ or /tʃi/ and [ta] can be go back to either /Ta/ or /ta/. This indeterminism, however, allows for smooth transitions during secondary splits. For example, unlike the derivational approach above, the introduction of /ti/ does not require an extra assumption about the lexicon or the re-analysis of existing syllables. The emergence of /ti/ leaves the /Ti/ → [tʃi] mapping intact, as shown in (34).

(34) The development of the /t/~/tʃ/ contrast in the context of [a] and [i]

1. single phoneme

	[t]
_a	/Ta/ → [ta]
_i	/Ti/ → [ti]

2. allophones

[t]	[tʃ]
/Ta/ → [ta]	
	/Ti/ → [tʃi]

15 See Lahiri and Reetz (2002) for arguments taking coronality as the default place feature.

3. partial opposition

[t]	[tʃ]
/Ta/ → [ta]	/tʃa/ → [tʃa]
	/Ti/ → [tʃi] /tʃi/ → [tʃi]

4. full opposition

[t]	[tʃ]
/Ta/ → [ta] /ta/ → [ta]	/tʃa/ → [tʃa]
/ti/ → [ti]	/Ti/ → [tʃi] /tʃi/ → [tʃi]

The underspecified approach also offers an intriguing explanation for the fact that the assimilation-triggering context is the most difficult to occupy. The explanation is based on the assumption that the birth of new underlying segments is triggered by a discrepancy between the auditory input and the stored perceptual representation. For example, matching [sa] and [ʃa] against underlying /Sa/ is possible, because /S/ is unspecified for the place of articulation, but the palatality of [ʃ] remains unmatched in the auditory input. The emerging /ʃ/ capitalizes on this dangling feature and offers a better match for auditory [ʃa]. In case of [si] and [ʃi], however, the extra palatality (or unary **I** feature) can be associated with the vowel part of underlying /Si/, and thus there is no perceptual evidence which a new phoneme can be based upon.[16] This explanation implies that for Japanese listeners, an auditory feature can match either the consonant or the vowel part of a CV sequence.[17] It is possible if the mora is assumed to be a functional unit for phonological encoding that can accommodate place features (Kureta, Fushimi, and Tatsumi 2006; Coleman 1998; Otake, this volume). The reason why consonant contrasts can still occupy the disadvantageous contexts is that listeners are capable of abstracting away linguistic content and focus merely on the acoustic signal (Mattingly et al. 1971). More salient differences in the signal make this process easier. This is a point where the representational approach can interface with phonetic explanations.

5.4 Summary

This brief overview has not listed all possible approaches to the representation of Japanese consonants, nor has it evaluated them thoroughly. It only outlined some

16 This explanation is a drastic simplification. A more comprehensive analysis can be worked out using features associated with moraic units (Declarative Phonology: Coleman 1998), and underspecified perception (Lahiri and Reetz 2002).

17 Consequently, it implies that speakers who are successful at segmenting speech sounds into phonemes, as opposed to segmenting them into moraic units, are better at perceiving consonant contrasts in the context of high vowels. This hypothesis needs experimental confirmation (see Otake, this volume, for related discussion).

of the coordinates along which most of the representational alternatives can be positioned. While derivational approaches seem to enjoy the most attention within these coordinates (e.g., Hattori 1960; Shibatani 1990; Tsujimura 2007; Frellesvig 2010; Labrune 2012), underspecification approaches display viable alternatives. Yoshida's (1996) analysis in governmental phonology, Coleman's (1998) declarative phonology approach, or Akamatsu's (2000) functional views provide great starting points for investigations that are based on underspecification.

Although this chapter aims to show that asymmetries in the CV restrictions are related to articulatory and perceptual forces, it does not exclude the possibility that they also have cognitive correlates. Underlying representations, especially in connection with theories of speech perception (e.g., Lahiri and Reetz 2002) can help understand why certain consonant contrasts categorically avoid some vocalic environment in Japanese.

6 Conclusion

The historical development of Japanese consonant contrasts is a complex topic at the intersection of historical linguistics, phonetics, and phonology. It is impossible to cover the topic exhaustively within a single chapter. The goal of this study was only to highlight some of the common patterns in the phonotactics of consonants, and show that they are results of recurrent sound changes. Restricting the observations to sibilant fricatives, non-sibilant fricatives, and the stop/affricate systems, it was shown that articulatory and perceptual forces are responsible for these recurrent patterns. The first step towards consonantal contrasts usually involves an articulatorily-motivated change resulting in positional allophones in the context of high vowels (e.g., [ʃi], [tʃi], [tsu], [çi], [ɸu]). Next, the distribution of allophones extends to other vocalic environments, most often under the pressure of loanwords. The expansions of consonant contrasts are hindered exactly in those vocalic environments where the allophony took place in the first place (e.g., /ʃi/ ↔ /si/, /tu/ ↔ /tsu/). The difficulty to accommodate contrasts in these environments has perceptual bases. The high vowels provide the least salient perceptual cues for the contrasts. The overall development of Japanese consonants seems to agree with the general observation from historical linguistics that sound changes are perception-oriented even though the seeds for change may have articulatory origins (Hyman 1976: 416).

After overviewing the history of the three consonant groups, the topic of underlying representation was addressed. The role of this theoretical addition to the otherwise phonetically inclined study was to investigate the possibility that phonotactic limitations can be traced back to cognitive origins. The way Japanese speakers store and manipulate sounds may be responsible for the most severe CV restrictions.

It was found that theories using underspecified segments present a promising direction to model emerging contrasts and explain phonotactic gaps. The thorough inspection of this direction is left for future research.

Some additional topics which were beyond the scope of this study include the analysis of nasals (i.e., [n] ↔ [ɲ]), voiced obstruents (e.g., [t] ↔ [d], [z] ↔ [ʒ]) and glides (i.e., /jV/ ↔ /V/ and /wV/ ↔ /V/). A brief look at the syllabary suggests that similar recurrent patterns centering on high vowels are to be found. The contextual effect of formant transitions is expected to be more pronounced with glides (e.g., /i/ = /ji/, /e/ = /je/; /u/ = /wu/, /o/ = /wo/) as perceptual oppositions cannot rely much on other consonantal cues.

The historical process of how consonant allophones were born and penetrated the phonotactic landscape of Japanese is one of the most intriguing problems in Japanese segmental phonology. The ongoing changes in the consonant system of contemporary Japanese makes the topic a promising field for historical, phonetic, and phonological studies. Hopefully this overview, while only scratching the surface, managed to raise some thought-provoking ideas for future studies.

Acknowledgments

I appreciate deep knowledge and insightful comments of Donna Erickson, Tomoaki Takayama, Timothy Vance, and Haruo Kubozono whose views greatly influenced the contents of this article. I also want to thank members of PAIK for providing invaluable feedback on earlier versions of the research.

References

Akamatsu, Tsutomu. 2000. *Japanese phonology: A functional approach*. München: Lincom Europa.

Anderson, Stephen R. 1985. *Phonology in the twentieth century*. Chicago and London: The University of Chicago Press.

Anttila, Ramino. 1972. *An introduction to historical and comparative linguistics*. New York: Macmillan.

Arakawa, Sobei. 1932. *Gairaigogaku yōsetsu* [An introduction to the study of loan-words]. Nagoya: Arakawa Sōbei.

Arisaka, Hideyo. 1944. *Kokugo on'inshi no kenkyū* [Studies of phonological history of Japanese]. Tokyo: Meiseidō Shoten.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge, New York: Cambridge University Press.

Blevins, James P. 2013. American descriptivism ('structuralism'). In Keith Allan (ed.), *The Oxford handbook of the history of linguistic*, 419–438. Oxford: Oxford University Press.

Bloch, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26. 86–125.

de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Boersma, Paul. 1998. *Functional phonology. Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bynon, Theodora. 1977. *Historical linguistics*. New York: Cambridge University Press.
Campbell, Lyle. 1998. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
Clark, John and Colin Yallop. 1990. *An introduction to phonetics and phonology*. Oxford, U.K. and Cambridge, MA: Blackwell.
Coleman, John. 1998. *Phonological representations: Their names, forms and powers*. Cambridge: Cambridge University Press.
Crystal, David. 1980. *A dictionary of linguistics and phonetics*. Oxford: Blackwell.
Dohlus, Katrin. 2004. The adaptation of German front rounded vowels into Japanese. *Kobe Papers in Linguistics* 4. 1–19.
Flemming, Edward. 1995. *Auditory representation in phonology*. Los Angeles: University of California dissertation.
Foulkes, Paul. 1997. Historical laboratory phonology – Investigating /p/ > /f/ > /h/ changes. *Language and Speech* 40(3). 249–276.
Frellesvig, Bjarke. 2010. *A history of the Japanese language*. New York: Cambridge University Press.
Fujisaki, Hiroya and Osamu Kunisaki. 1978. Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(1). 21–27.
Funatsu, Seiya and Shigeru Kiritani. 2000. Effect of following vowel on perception of second language fricatives. *Journal of the Phonetic Society of Japan* 4(2). 72–80.
Gordon, Matthew, Paul Barthmaier and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32(2). 141–174.
Hall, Tracy A. 1995. Inalterability and sibilant neutralization in Gujarati. *Lingua* 98. 11–20.
Hall, Tracy A, Silke Hamann and Zygis Marzena. 2006. The phonetic motivation for phonological assibilation. *Journal of the International Phonetic Association* 36(1). 59–81.
Hamada, Atsushi. 1952. Chosen ban “Irofa” genbun [Original Korean version of Irofa]. *Kokugo Kokubun* 21. 77–90.
Hamano, Shoko. 2000. Voicing of obstruents in Old Japanese: Evidence from the sound-symbolic stratum. *Journal of East Asian Linguistics* 9. 207–225.
Harris, Katherine S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1. 1–7.
Hashimoto, Shinkichi. 1950. *Kokugo on’in no kenkyū* [Studies on Japanese phonology]. Tokyo: Iwanami.
Hattori, Shiro. 1960. *Gengogaku no hōhō* [Methodologies in linguistics]. Tokyo: Iwanami.
Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62. 321–351.
Hirai, Sawako, Keichi Yasu, Takayuki Arai and Kyoko Iitaka. 2005. Acoustic cues in fricative perception for Japanese native speakers. *IEICE technical report. Speech* 104. 25–30.
Hoenigswald, Henry M. 1960. *Language change and linguistic reconstruction*. London: The University of Chicago Press.
Hyman, Larry M. 1976. Phonologization. In Alphonse Juilland, Andrew M. Devine and Laurence D. Stephens (eds.), *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, 407–418. Saratoga, California: Anma Libri.
Inkelas, Sharon and Young-mee Yu Cho. 1993. Inalterability as prespecification. *Language* 69(3). 529–574.
Irwin, Mark. 2011. *Loanwords in Japanese*. Amsterdam and Philadelphia: John Benjamins.
Ishiwata, Toshio. 2001. *Gairaigo no sōgōteki kenkyū* [A comprehensive research of loanwords]. Tokyo: Tōkyōdō Shuppan.
Ito, Junko and Armin Mester. 1995. Japanese Phonology. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Cambridge, MA: Blackwell.
Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Jaeger, Jeri J. 1978. Speech aerodynamics and phonological universals. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society 4*, 311–329.
Janson, Tore. 1986. Cross-linguistic trends in the frequency of CV sequences. *Phonology Yearbook* 3. 179–195.
Jōo, Hakutarō. 1982. *Onseigaku* [Phonetics]. Tokyo: Aporon Ongaku Kōgyō.
Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kawai, Junji. 2003. Markedness in casual speech. *Japanese Linguistics* 14. 73–97.
Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to the sounds of Japanese]. Tokyo: Ōfūsha.
Kawasaki, Haruko. 1982. *An acoustical basis for universal constraints on sound sequences*. Berkeley: University of California dissertation.
Kindaichi, Haruhiko. 1932. *Kokugo on'inron* [Japanese phonology]. Tokyo: Katae Shoin.
Kiparsky, Paul. 1968. Linguistic universals and linguistic change. Reprinted in Paul Kiparsky 1982. *Explanation in phonology*. Dordrecht: Foris.
Kobayashi, Akemi. 1981. Ennin no kijutsu suru sansukuritto onsetsu 'ca' no onka: Kyū-seiki nihongo suitei no kokoromi [The phonetic value of the Sanskrit 'ca' as described by Ennin: An attempt to reconstruct Japanese in the 9th century]. *Osaka Gaikokugo Daigaku Gakuhō* 52. 63–80.
Koizumi, Tamotsu. 1996. *Onseigaku nyūmon* [An introduction to phonetics]. Tokyo: Daigaku Shorin.
Krämer, Martin. 2012. *Underlying representations.* Cambridge: Cambridge University Press.
Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kureta, Yoichi, Takao Fushimi and Itaru F. Tatsumi. 2006. The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(5). 1102–1119.
Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.
Ladefoged, Peter and Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford: Blackwell.
Lahiri, Aditi and Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 637–675. Berlin: Mouton de Gruyter.
Lambacher, Stephen, Williams Martens, Brian Nelson and Jonathan Berman. 2001. Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *Acoustical Science and Technology* 22(5). 334–343.
Lange, Roland A. 1973. *The phonology of eight-century Japanese: A reconstruction based upon written records.* Tokyo: Sophia University Press.
Lass, Roger. 1984. *Phonology: An introduction to basic concepts.* Cambridge: Cambridge University Press.
Laver, John. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.
Liljencrants, Johan and Bjorn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4). 839–862.
Lindblom, Bjorn. 1983. Economy of speech gestures. In Peter MacNeilage (ed.), *The production of speech*, 217–245. New York: Springer-Verlag.
Lisker, Leigh. 1986. Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29. 3–11.
Mabuchi, Kazuo. 1971. *Kokugo on'inron* [Japanese phonology]. Tokyo: Kasama Shoin.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
Mann, Virginia A. and Bruno H. Repp. 1980. Influence of vocalic context on perception of the [sh]-[s] distinction. *Perception and Psychophysics* 28(3). 213–228.
Martin, Samuel E. 1987. *The Japanese language through time*. New Haven: Yale University Press.
Matsui, Michinao F. 2012. Shakuyōgo ni okeru sokuon seiki no yokusei yōin [An inhibitory factor of occurrence of voiceless geminate consonants in Japanese loanwords]. *Theoretical and Applied Linguistics at Kobe Shoin* 15. 49–102.
Mattingly, Ignatus G, Alvin M. Liberman, Ann K. Syrdal and Terry Halwes. 1971. Discrimination in speech and nonspeech modes. *Cognitive Psychology* 2(2). 131–157.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
McMahon, April M. S. 2000. *Change, chance, and optimality*. New York: Oxford University Press.
Miller, Roy A. 1967. *The Japanese language*. Chicago: University of Chicago Press.
Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Nishida, Naotoshi. 2001. *Nihongoshi ronkō* [Investigations on the history of Japanese]. Osaka: Izumi Shoin.
Nishida, Mariko. 2010. Nihongo no tagyō shiin onso ni kanshite [A study of Ta-line consonant phonemes in Japanese]. *Journal of Sapporo International University* 41. 149–155.
Nowak, Pawel M. 2006. The role of vowel transition and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 34. 139–152.
Ohala, John J. 1974. Experimental historical phonology. In John Mathieson Anderson and Charles Jones (eds.), *Historical linguistics II. Theory and description in phonology*, 353–389. Amsterdam: North Holland.
Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick and Mary F. Miller (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 17: Papers from the Parasession on Language and Behavior*, 178–203.
Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Padgett, Jaye. 2001. Contrast dispersion and Russian palatalization. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*, 187–218. Tokyo: Academic Press.
Paradis, Carole and Jean-François Prunet 2000. Nasal vowels as two segments: Evidence from borrowings. *Language* 76. 324–357.
Pintér, Gábor. 2005. Where have the Japanese velar glides gone? Paper presented at LingO – The Oxford Postgraduate Linguistics Conference, Mansfield College, 23–24 September.
Pintér, Gábor. 2007. The interaction of contextual vowels and formant transitions in Japanese fricative perception. Talk presented at the 315th regular meeting of the Phonetic Society of Japan, The University of Tokyo, 30 June.
Raphael, Lawrence J. 2005. Acoustic cues to the perception of segmental phonemes. In David B. Pisoni and Robert E. Remez (eds.), *The handbook of speech recognition*, 182–206. Malden, Massachusetts: Blackwell.
Rodriguez, Joao. 1955 [1604]. *Nihon daibunten* [A Japanese translation of Rodriguez 1604–1608]. Tokyo: Sanseido.
Ségéral, Philippe and Tobias Scheer. 2001. Abstractness in phonology: The case of virtual geminates. In Katarzyna Dziubalska Kolaczyk (ed.), *Constraints and preferences*, 311–337. New York: Mouton de Gruyter.
Shaw, Jason and Rahul Balusu. 2010. Language contact and phonological contrast: The case of coronal affricates in Japanese loans. In Muriel Norde, Bob De Jonge and Cornelius Hasselblatt (eds.), *Language contact: New perspectives*, 155–180. Amsterdam: John Benjamins.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Steriade, Donca. 1988. Review of Clements and Keyser 1983. CV phonology. *Language* 64. 118–129.

Steriade, Donca. 2001. *The phonology of perceptibility effects: The P-map and its consequences for constraint organization*. Unpublished manuscript.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Tews, Andrea. 2008. Japanese geminate perception in nonsense words involving German [f] and [x]. *Gengo kenkyu* 133. 133–145.

Takeyasu, Hajime. 2009. Masatsuon no sokuon chikaku ni okeru masatsu shūhasū tokusei no eikyō [Effects of the spectral properties of frication on perception of singleton/geminate fricatives]. *On'in Kenkyū* [Phonological Studies] 12. 31–38.

Tsujimura, Natsuko. 2007 [1996]. *An introduction to Japanese linguistics*. Cambridge, MA and Oxford: Blackwell.

Trón, Viktor and Péter, Rebrus. 2001. Morphology and the hierarchical lexicon. *Acta Linguistica Hungarica 48*. 101–136.

Umegaki, Minoru. 1944. *Nihon gairaigo no kenkyū* [A study on loanwords in Japanese]. Osaka: Seinen Tsūshinsha.

Unger, James M. 1993 [1975]. *Studies in early Japanese morphophonemics*. Indiana University Linguistics Club Publications.

Unger, James M. 2004. Alternations of m and b in Early Middle Japanese: The deeper significance of the sound-symbolic stratum. *Japan Language and Literature* 38. 323–337.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Watanabe, Minoru. 1997. *Nihongoshi yōsetsu* [An overview of the history of Japanese]. Tokyo: Iwanami.

Whalen, Douglas H. 1981. Effects of vocalic formant transition and vowel quality on the English [s]-[sh] boundary. *Journal of the Acoustical Society of America* 69. 275–282.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Yamazaki, Shūichi, Akira Utsugi and Hsin-ying Pan. 2004. An acoustic phonetic study of voiceless alveolo-palatal fricatives in Japanese, Korean, and Chinese. *Ippan Gengogaku Ronsō* [Journal of General Linguistics] 7. 1–27.

Yoshida, Shōhei. 1996. *Phonological government in Japanese*. Canberra: The Australian National University dissertation.

Zipf, George K. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley Press.

Masako Fujimoto

4 Vowel devoicing

1 Introduction

This chapter describes various aspects of vowel devoicing phenomena in Japanese. Vowel devoicing (hereafter devoicing) appears in many dialects of Japanese. For example, the vowels /i/ in *sika* 'deer' and /u/ in *kusa* 'grass' often lose voicing. Traditional literature states that high vowels /i/ and /u/ undergo devoicing in the moras /pi/, /pu/, /ki/, /ku/, /si/, /sju/, /ti/, /tju/, /hi/, /su/, /tu/, and /hu/ at normal speech rate, especially when these moras are followed by a mora with a voiceless consonant or by silence (Kawakami 1977; Amanuma, Ōtsubo, and Mizutani 1993). More generally, the high vowels are devoiced if they are placed between voiceless obstruents, or a voiceless obstruent and a pause (phrase break) (Sakuma 1929; Han 1962; McCawley 1968; Ōso 1973; Nihon Onsei Gakkai 1976; Hirayama 1985; Vance 1987; Sibata 1988; Hibiya 1999; Tanaka and Kubozono 1999). These contexts are formulated as follows.

(1) V → V̥ / C̥ _ C̥
[+high] [-voice] [-voice]

(2) V → V̥ / C̥ _ #
[+high] [-voice]

Devoicing is often claimed to be obligatory in Tokyo Japanese (Sakuma 1959; Hirayama 1985). If vowels in these conditions are uttered voiced, they sound unnatural (Sakuma 1933; Hirayama 1985). According to Hirayama (1985), devoicing functions to delineate words or phrases and to enhance the crispness of pronunciation. Since moderate use of devoicing is highly advocated in standard Japanese (Sakuma 1959; Hirayama 1985), devoicing is included as one of the topics in broadcasting textbooks (e.g. NHK 2005). Authorized leading dictionaries of pronunciation and accent such as NHK (1985) and Kindaichi and Akinaga (2001) mark the moras that are subject to vowel devoicing.

Contrary to the description that devoicing is obligatory in Tokyo Japanese, its actual occurrence diminishes due to many factors such as consonantal environment, accent, speech rate, and dialects. Thus, the formulations shown above serve only as first approximations. Devoicing in Japanese has been extensively studied in acoustic and in articulatory aspects. Major findings from these studies are summarized in section 2, according to segmental, suprasegmental, and sociolinguistic factors. Perceptual (and psycholinguistic) studies are briefly sketched in section 3. In section 4, other topics related to vowel devoicing are overviewed.

Devoiced and voiced vowels are allophones since the variation does not alter the meaning of the word. Hence, it is by no means phonemic/phonological in a narrow sense. However, since devoicing is systematic in standard Tokyo Japanese and since it readily occurs in certain conditions, many researchers treat devoicing within a phonological framework. In this chapter, devoicing is discussed mainly from a phonetic point of view with occasional mention of its phonological treatment.

Some scholars distinguish between devoiced and deleted vowels. This issue is discussed in 4.4. In the rest of this chapter the term devoicing is used unless otherwise noted. In the following sections, symbols C, C_1, C_2, C̥, C̬, V, V_h, V_{nh}, #, and Q are used to denote consonants, the preceding consonants, the following consonants, voiceless consonants, voiced consonants, vowels, high vowels, non-high vowels, a pause (phrase break), and the first half of a geminate consonant, respectively, as in $C_1V_{nh}C_2$ or $C_1V_hQC_2$. The abbreviations St, Af, and Fr stand for voiceless stops, voiceless affricates, voiceless fricatives, respectively. St/Af-Fr, for example, indicates that the preceding consonant (C_1) is a stop or an affricate, while the following consonant (C_2) is a fricative. Similarly, Af/Fr-QSt denotes that the preceding consonant (C_1) is an affricate or a fricative and the following consonant (C_2) is a geminate stop.

2 Factors that affect devoicing

2.1 Vowels

2.1.1 High vowels

Devoicing is most frequently observed in high vowels /i/ and /u/. It is presumably because they are shorter than non-high vowels (Sakuma 1929). Quantitative studies confirmed that /i/ and /u/ are shorter than /a/, /e/, and /o/ in Japanese (Hiki, Kanamori, and Oizumi 1967; Sagisaka and Tohkura 1984). In Sagisaka and Tohkura (1984), the durations of the vowels /a/, /e/, /o/, /i/, /u/ uttered in short phrases are 86ms, 79ms, 71ms, 61ms, and 58ms, respectively. Among the two high vowels, /u/ is reportedly shorter (Han 1962; Hiki, Kanamori, and Oizumi 1967; Sagisaka and Tohkura 1984). High vowels are intrinsically shorter than non-high vowels (Lehiste 1977). Vance (1987) infers that the durational difference between high and non-high vowels is greater in Japanese than other languages such as Swedish (Elert 1964, cited in Lehiste 1977).

On the frequency of the devoicing for /i/ and /u/, studies disagree. Maekawa (1983) and Yoshida (2002) found no difference between the two vowels. Imai (2004) reported that /i/ is more frequently devoiced than /u/. On the other hand, Han (1962) and Imai (2010) showed that /u/ is more frequently devoiced than /i/. Also, in the Corpus of Spontaneous Japanese (2004) (hereafter CSJ corpus; see Maekawa, this volume, for more details about corpus-based studies), which involves more than 300,000 vowels, /u/ is more frequently devoiced than /i/ when only devoicing

between voiceless consonants is counted: the devoicing rate is 17.37% for /i/ and 20.91% for /u/ (Maekawa and Kikuchi 2005).

2.1.2 Non-high vowels and long vowels

Like high vowels, non-high vowels are subject to devoicing, too, if they are surrounded by voiceless obstruents. For example, the first vowels often devoice in the words *kakaru* 'to hang on', *kokoro* 'heart' (Sakuma 1929), *katana* 'sword', *kakou* 'to enclose', *kome* 'rice', *kakkoo* 'appearance, downhill' (Kawakami 1977), *kakasi* 'scarecrow' (Sakurai 1985), *kesyoo* 'make-up' and *ketatamasii* 'noisy' (Amanuma, Ōtsubo, and Mizutani 1993). The Japanese Language Council (1954) regards the devoicing in this environment as non-standard and a dialectal variant (Amanuma, Ōtsubo, and Mizutani 1993). Kawakami (1977) mentions that non-high vowel devoicing is optional and varies interpersonally among Tokyo speakers.

Acoustic studies have shown that devoicing in non-high vowels is indeed non-systematic. In a production study using non-words, devoicing of /a/ occurred in /ha/ before a voiceless consonant only at fast speech, and in /sa/ before voiced consonants at nearly maximum speed (Maekawa 1990). In the speech data that involved eleven male and female announcers/narrators, non-high vowels devoiced only eleven times out of over 38,000 tokens: there are ten instances of devoiced /a/ (e.g. [kḁkɯte:] 'local train', [hanayakḁ] 'brilliant') and one instance of /o/ ([te:ʃo̥kɯ] 'steady job'), but no instance of /e/ devoicing (Kawai et al. 1995). In the CSJ corpus, non-high vowel devoicing between voiceless consonants rarely occurred: the rates of devoicing in /a/, /e/, and /o/ are 2.10%, 3.31%, and 3.45%, respectively (Maekawa and Kikuchi 2005). In the telephone corpus, devoicing is reported in /a/ in [kḁgoʃima] 'Kagoshima (city name)' and [kḁkarɯ] 'to take', in /o/ in [ko̥ko] 'here' and in /e/ in [he̥ta] 'clumsy', although the devoicing rates are not provided (Komatsu and Aoyagi 2005; Arai, Warner, and Greenberg 2007).

According to Sakuma (1929), devoicing is least likely to occur in /e/, although the first /e/'s in *tesuki* 'be not busy', *kessite* 'never', and *sekkaku* 'at great pains' are devoiced. Kawai et al. (1995) agree with Sakuma's (1929) observation. However, devoicing is more frequent in /e/ than /a/ in the CSJ corpus cited above. Hence, it is not clear which of the three non-high vowels is most likely to devoice.

Interestingly, non-high vowels tend to devoice if identical moras come next to each other (Sakuma 1959; Sakurai 1985). As mentioned above, the devoicing rate of non-high vowels between voiceless consonants is low in the CSJ corpus. However, when limited to the first vowel of two identical moras, the rate increases to 10.5%, 4.3%, and 22.3% for /a/, /e/, and /o/, respectively (Maekawa and Kikuchi 2005). The manner of adjacent consonants does not play a crucial role in the occurrence of the non-high vowel devoicing, whereas it does in high vowel devoicing (Maekawa and Kikuchi 2005) (see section 2.2 on consonantal conditions). These observations

suggest that the mechanism of devoicing may differ between high and non-high vowels. Studies on non-high vowel devoicing are still limited. Detailed examinations are desirable.

Devoicing of long vowels does not generally occur. It is plausibly because long vowels are relatively longer in duration than single vowels. In the CSJ corpus, long vowels are devoiced only 49 times, which is a negligibly small fraction out of 300,000 vowels in the corpus (Maekawa and Kikuchi 2005).

2.1.3 Vowel height of the following mora

Devoicing is claimed to be more frequent if the vowel in the immediately following mora is a non-high vowel than if it is a high vowel (Inoue 1968; Maekawa 1989). Acoustic studies show that devoicing is more frequent when the vowel of the following mora is /a/ than /u/ (Maekawa 1990; Yoshida 2002; Byun 2007). As discussed in section 2.1.2, non-high vowel devoicing occurs relatively frequently when identical moras come next to each other (e.g. *kakasi* [kḁkaʃi] 'scarecrow' and *kokoro* [ko̥koro] 'heart'). In this condition, the vowel of the following mora is a non-high vowel. On the other hand, high vowel devoicing is less frequent when identical moras come next to each other as in *tutumu* 'to wrap' and *kukuru* 'to bundle' (Hino 1966). In these two conditions, the vowel height of the following mora differs: non-high in the former case and high in the latter. These examples may also suggest that non-high vowels facilitate devoicing in the preceding moras. An appropriate explanation for this remains unknown. Detailed and quantitative examinations are called for on this issue.

2.2 Consonantal conditions

2.2.1 Acoustic studies

Devoicing is said to occur when high vowels are placed between voiceless obstruents. However, actual occurrence of devoicing differs significantly depending on the consonantal condition. Qualitative and quantitative studies agree that the manner of articulation of the preceding and/or the following consonants strongly affects the devoicing rate, but their interactions are seemingly complicated. Some studies claim that the preceding consonants have a stronger effect (e.g. Han 1962), whereas others emphasize the effect of the following consonants (e.g. Byun 2007). Devoicing is more frequent when a fricative precedes the vowel than when a stop does (Han 1962; Maekawa 1983, 1989; Takeda and Kuwabara 1987; Sugito 1996; Hashimoto et al. 1997; Kondo 1997; Imai 2004). On the other hand, devoicing is less frequent when a fricative follows the vowel than when a stop does (Sakurai 1985; Takeda

and Kuwabara 1987; Yoshida and Sagisaka 1990; Sugito 1996; Imai 2004; Maekawa and Kikuchi 2005). Namely, fricatives facilitate devoicing when they precede a vowel, but suppress it when they follow a vowel. Moreover, devoicing is less frequent when the preceding and following consonants are both fricatives (Sakuma 1929; Sakurai 1966; Kimura, Kaiki, and Kitoh 1998; Yoshida 2002; Fujimoto and Kiritani 2003; Maekawa and Kikuchi 2005; Fujimoto 2005). Among fricatives, /h/ suppresses devoicing more than /s/ does in the post-vocalic position (Nagano-Madsen 1994b; Fujimoto and Kiritani 2003; Fujimoto 2004a), and [ʃ] suppresses devoicing more than [s] (Nagano-Madsen 1995).

Given the asymmetric effects found in fricatives, the combination of preceding and following consonants better estimates devoicing probability. By analyzing speech data of a male announcer, Kimura, Kaiki, and Kitoh (1998) ranked the devoicing rates as shown below. Note that voiceless affricates are limited to [tʃ] and [ts] in Japanese.

most frequently devoiced:	Af/Fr – St/Af
moderately devoiced:	St – St/Af/Fr
frequently voiced:	Af – Fr, Fr_1 – Fr_2 (sequence of different fricatives)
seldom devoiced:	Fr_1 – Fr_1(sequence of same fricatives)

According to the analysis of the CSJ corpus (Maekawa and Kikuchi 2005), the devoicing rate was highest when a fricative is followed by a stop (Fr-St), and second highest when a fricative is followed by an affricate (Fr-Af). In contrast, the rate was lowest when an affricate is followed by a fricative (Af-Fr) and second lowest when a fricative is followed by a fricative (Fr-Fr). Moreover, the devoicing rate was highest when C_2 was a stop and lowest when C_2 was a fricative. In Maekawa and Kikuchi's data, the devoicing rate is generally lower when /h/ is at the C_2 position than when /s/ is at the same position.

The results of Kimura, Kaiki, and Kitoh (1998) and Maekawa and Kikuchi (2005) suggest that affricates behave similar to fricatives when they precede a vowel and to stops when they follow a vowel. This view is reasonable since an affricate is "a stop followed by a homorganic fricative" (Ladefoged and Johnson 2011: 67). At C1 position, affricates show less frequent devoicing than fricatives do (Han 1962), a more frequent rate than fricatives (Yoshida and Sagisaka 1990) and a similar rate as fricatives (Takeda and Kuwabara 1987; Fujimoto and Kiritani 2003). The disagreement of these studies is possibly due to other factors such as consonants in the following moras. Affricates in C2 position yield the condition of consecutive devoicing, since they are limited to [tʃi] and [tsɯ] in Japanese. This may reduce the probability of devoicing of the preceding vowel.

Most of the studies mentioned above used real words, where many conditions such as phonological environments, word duration and accent are difficult to control. The results can be skewed depending on the materials contained in the data

set. This may cause the disagreement among studies as to, for example, the importance of the preceding or the following consonants. The consonantal effects appear rather explicitly in the studies using non-words. Table 1 shows the average devoicing rate of /i/ of ten Tokyo speakers in unaccented /C_1iC_2e/ non-words, where /k, (t), s, h/ are systematically combined (Fujimoto 2004a). Note that /s/ and /h/ are [ʃ] and [ç] in C_1 position and [s] and [h] in C_2, respectively. The result generally agrees with that of Kimura, Kaiki, and Kitoh (1998). Devoicing rate is almost 100% when C_1-C_2 is either 'St-St', 'St-Fr' or 'Fr-St', if /h/ at C_2 position is excluded. The rate drops to 43% on average, when the combination is 'Fr-Fr' (/sise/, /sihe/, /hise/, and /hihe/). Among 'Fr-Fr' combinations, devoicing rate of the same consonants (/sise/ and /hihe/) is not necessarily lower than that of different consonants (/sihe/ and /hise/), which differs from Kimura, Kaiki, and Kitoh's (1998) result. However, the critical factor that suppresses devoicing is not the combination of the two consonants but the /h/ in C_2 position. The rate is as low as 17% on average when the C_2 is /h/ regardless of the C_1 (/kihe/, /sihe/, and /hihe/).

Table 1: Devoicing rate of /i/ in unaccented /C_1iC_2e/ non-words uttered in a frame sentence. (Adapted from Fujimoto 2004a)

C_1 \ C_2	k	t	s	h
k	100%	100%	100%	23%
s	100%	98%	58%	13%
h	100%	100%	85%	15%

From these results, it is reasonable to categorize the consonantal conditions of high vowel devoicing into two types: typical and atypical. In the 'typical' consonantal conditions devoicing occurs systematically and regularly in Tokyo speakers as described by many traditional studies. They are 'St-St', 'St-Fr (except for /h/)' and 'Fr-St'. In the 'atypical' consonantal conditions, devoicing occurs randomly with greater inter-speaker variation. They are 'Fr-Fr' and 'St/Af/Fr-/h/'. Within atypical consonantal conditions, probability of devoicing is lowest when the target vowel is followed by an /h/.

As was shown above, a consonantal combination of Fr-Fr and Af-Fr and /h/ in the following position suppresses devoicing both in experimentally-controlled studies (Yoshida 2002; Fujimoto and Kiritani 2003; Fujimoto 2004a) and spontaneous speech (Maekawa and Kikuchi 2005). Kawatsu and Maekawa (2009) and Maekawa (2011) surveyed the correlation between the devoicing rate and the cepstrum distance of the preceding and the following consonants, and found that the devoicing rate is low when the cepstrum distance is small as in Fr-Fr, Af-Fr. Smaller cepstrum distance means that two consonants sound similar. This suggests that speakers suppress devoicing so as to avoid perceptual confusion. They also found that the devoicing rate is very low when /h/ or /hj/ follows a vowel regardless of the preced-

ing consonants, although their cepstrum distance is large (Kawatsu and Maekawa 2009; Maekawa 2011). They attribute this smaller devoicing rate to the voicing tendency of these consonants /h/ and /hj/ (Kawatsu and Maekawa 2009; Maekawa 2011).

2.2.2 Physiological studies

Voicing is primarily controlled by glottal vibration. Hence, it is ideal to examine the glottal state during /CVC/. In this respect, the devoicing phenomenon in Japanese has been intensively studied through physiological examinations. Conventional filming and high-speed video recordings (see, for example, Kiritani, Imagawa, and Hirose 1996) enable the direct visual inspection of vocal folds. Photoglottography (hereafter PGG), or transillumination (Lisker et al. 1969), records the amount of light that passes though the glottis while it opens and closes during speech (Hirose 1999). Due to technical difficulties and ethical limitations, the subjects examined in these studies are limited. Also the vowels used in the filming and PGG are usually limited to front vowels /i/ and /e/, since, during back vowels, the epiglottis tilts backward and often hides laryngeal views. Despite these limitations, previous physiological studies revealed the characteristics of glottal manifestation during devoicing, which acoustic analysis alone could not elucidate.

In speech production, the vocal folds generally adduct to vibrate during voiced segments and abduct during voiceless segments. Hence, the glottis is expected to show an 'open-close-open' pattern for /C̥VC̥/ sequence. However, strikingly, almost all studies agree that the glottis opens without any closing movement during /C̥VC̥/ sequence when the intermediate vowel is devoiced (Sawashima 1969, 1971b; Sawashima and Miyazaki 1973; Sawashima and Niimi 1974; Yoshioka 1981; Yoshioka, Löfqvist, and Hirose 1982; Fujimoto et al. 2002; Fujimoto 2004b). That is, the glottal openings show a single phase (i.e. mono-modal). This mono-modal pattern appears with no exception in typical consonantal conditions when Tokyo dialect speakers are concerned. Namely, the devoiced vowel is produced with an open glottis. This is consistent with Sakuma's (1929) intuition that devoiced vowels are produced not by vocal folds but by breath. Also, the degree of glottal opening for devoiced /C̥$_1$VC̥$_2$/ tends to be greater than that for each of the single consonants /C̥$_1$/ and /C̥$_2$/ (Sawashima 1971b; Sawashima and Miyazaki 1973; Sawashima and Niimi 1974; Fujimoto et al. 2002).

Figure 1 compares the glottal opening patterns for unaccented non-words, /kide/ and /kite/, as produced by a Tokyo speaker. Note that the figure includes /e/ in the preceding phrase of the frame sentence, as many other figures in the following sections do. The upper signal shows the speech wave and the lower, the PGG signal which corresponds to the glottal opening area. As can be seen in the speech wave, the vowel /i/ is voiced in /kide/ and devoiced in /kite/. In the PGG signal, the glottis opens for /k/ and closes during the voiced segments /ide/ in /kide/, as expected for /C̥VC̬V/ sequence. On the other hand, in /kite/, the glottal opening

shows a mono-modal pattern during /kit/ with no trace of glottal closure movement for the vowel. The degree of glottal opening during devoiced /kit/ is larger than that during a single /k/ in /kide/. An electromagnetic articulographic (EMA) and PGG study of another speaker showed that tongue movements are identical for /kide/ with voiced /i/ and /kite/ with devoiced /i/, while glottal opening is mono-modal during /kit/ (Funatsu and Fujimoto 2011). This suggests that devoicing of /i/ is accomplished solely by laryngeal articulation.

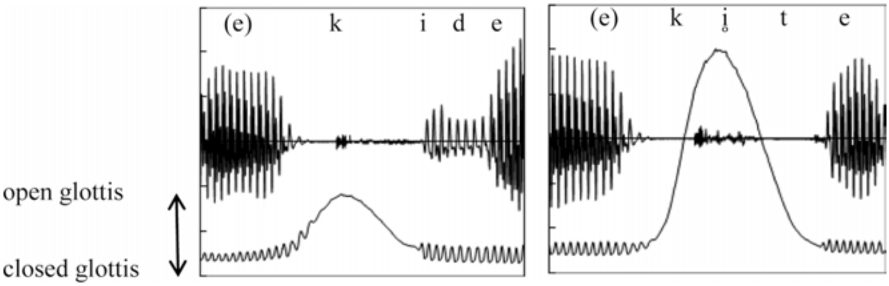


Figure 1: Speech wave (top) and glottographic signal (bottom) during the unaccented non-words /kide/ (left) and /kite/ (right) by a Tokyo speaker. Vowel /i/ is devoiced in /kite/. (Modified from Fujimoto et al. 2002)

Electromyographic data give further evidence for these observations. Generally, voiced and voiceless segments are reciprocally controlled by the Interarytenoid (INT), a glottal closing muscle, and the Posterior Cricoarytenoid (PCA), the only glottal opening muscle of the inner larynx (Borden and Harris 1980). Hence, PCA is expected to activate twice, for C_1 and C_2, during $/\underset{\circ}{C}_1V\underset{\circ}{C}_2/$. However, previous studies found that, when a vowel is devoiced, activation of PCA appears only once during the [C̥V̥C̥] with suppressed activation of INT (Hirose 1971; Sawashima, Hirose, and Yoshioka 1978; Yoshioka 1981). This suggests that the mono-modal glottal opening pattern with a devoiced vowel is executed by motor control at the myographic command or higher level (Sawashima, Hirose, and Yoshioka 1978; Yoshioka 1981; Yoshioka, Löfqvist, and Hirose 1982). Thus, we can infer that a devoiced vowel is not a by-product of glottal assimilation. Rather, this mono-modal glottal opening pattern can be viewed as reorganized from /C/+/C/ into /CV̥C/ in order to cooperate with producing devoiced vowels. In this respect, devoicing is an intentional or positively controlled event as far as Tokyo speakers are concerned, although this does not necessarily mean that the speakers consciously produce devoiced vowels.

Exceptionally, double phase (i.e. ‘bimodal’) and ‘plateau-like’ opening patterns were observed in the previous studies. These tokens are limited to the atypical consonantal conditions, where vowels are surrounded by two fricatives (Fr-Fr), or a consonant followed by a geminate one (C_1-QC_2) (Sawashima 1971b; Yoshioka, Löfqvist, and Hirose 1982; Tsuchida 1997; Fujimoto et al. 1998; Fujimoto, Funatsu, and Fujimoto

2012). Bimodal patterns can be achieved by the concatenation of two openings with a closing movement for a vowel in between. This indicates that the glottis opens for the second consonant on the way to closing for the intermediate vowel but failed to reach full closure. In these cases, vowels can be voiced or devoiced depending largely on how narrow the closure is. When the closing degree is not sufficient, vocal folds fail to start vibrating. Plateau-like opening is regarded as the state when the two openings are more closely combined leaving no trace of closing movement. These bimodal and plateau-like patterns suggest that devoicing can be executed not only by glottal reorganization but also by glottal assimilation in atypical consonantal conditions. This analysis is compatible with the infrequent, random occurrence of devoicing in Fr-Fr and Fr-QFr sequences (see section 2.2.3 for geminates).

Figure 2 shows variation of glottal opening patterns during Fr-Fr. The figure compares four repetitions of /sise/ produced by a Tokyo speaker. The vowel /i/ is devoiced in the first and the third tokens from the left, and voiced in the second and the fourth. The voiced tokens show a bimodal glottal opening pattern with smaller second openings than the first. Two devoiced tokens differ in the glottal openings, since the first shows a bimodal pattern while the third, a mono-modal. Namely, the glottal control of the first token is similar to that in the second, third, and fourth tokens, but the vocal folds failed to vibrate, or the vibration is too weak to be detected in the speech signal. This can be viewed as unintentional devoicing. As for the third token, where the opening is mono-modal and large, the glottal opening pattern is plausibly reorganized, and this devoicing can be intentional. Hence, this speaker produces both intentional and unintentional devoicing for this atypical consonantal condition.

The duration of devoiced /sis/ sequence is apparently shorter in the third token than the first, since the duration of each panel is the same (400ms). It has been argued that the durations of moras are shorter when the vowels are devoiced than when they are voiced (e.g. Han 1962; Beckman 1982; Han 1994). These examples clearly demonstrate that the duration of the devoiced mora depends largely on the glottal opening pattern, which cannot be detected from the acoustic signal. The durations of devoiced moras are shorter when glottal opening of /C̥VC̥/ is reorganized than when it is not.

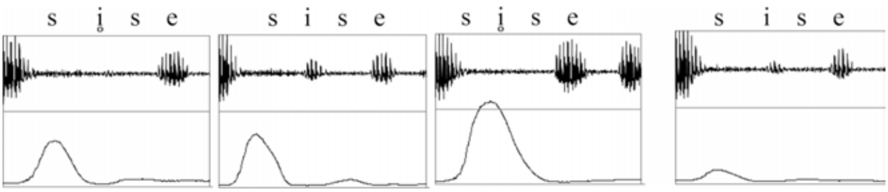


Figure 2: Speech wave (top) and glottographic signal (bottom) during the unaccented non-word /sise/ [ʃise] produced by a Tokyo speaker.

To sum, the devoicing pattern for Tokyo speaker is twofold: (i) categorical devoicing exemplified by typical consonantal conditions, and (ii) non-categorical devoicing exemplified by atypical consonantal conditions. The former may be neurologically controlled, and the glottal opening is reorganized into a mono-modal pattern. The latter is not necessarily neurologically controlled, and the glottal opening often shows a bimodal or plateau pattern.

In cases where /h/ follows a vowel, the fewer occurrences of devoicing can be basically attributed to supralaryngeal factors during the consonant, i.e. less constriction of the vocal tract. It has been shown that, although the degree of glottal opening is comparably large for both consonants /s/ and /h/, the vocal folds vibrate intervocalically throughout /h/ (Yoshioka, Löfqvist, and Hirose 1982). Hence, /h/ tends to be voiced while /s/ is not. This explains why /h/ in C_2 position suppresses devoicing. Although /h/ is generally categorized as a voiceless consonant, it has no voiced counterpart. This may accelerate the tendency of the voicing of /h/. When /h/ is voiced, the environment is no longer categorized as a general devoicing condition. Figure 3 compares four repetitions of /hihe/ produced by a Tokyo speaker (the same as in Figure 2). In the figure, the word-medial /h/ is realized as a voiced [ɦ] in all tokens, whereas the word-initial /h/ remains voiceless. Also, the vowel /i/ is voiced in all tokens while its intensity and the duration vary from token to token.

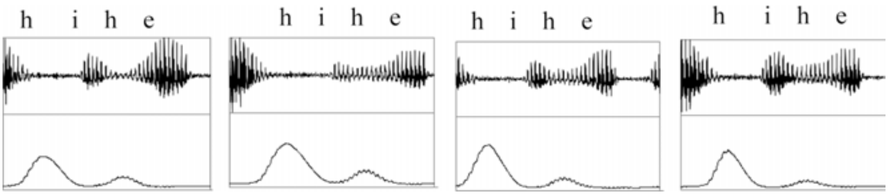


Figure 3: Speech wave (top) and glottographic signal (bottom) during the unaccented non-word /hihe/ produced by a Tokyo speaker.

2.2.3 Devoicing before geminate consonants

Devoicing of high vowels occurs before geminate voiceless consonants. Many studies treat single and geminate consonants equally in terms of devoicing (Nihon Onsei Gakkai 1976; Kawakami 1977; Amanuma, Ōtsubo, and Mizutani 1993). Block (1950) reports the devoicing of /i/ in *sissyoku* ‘unemployment,’ *sikken* ‘judgment,’ *sitta* ‘knew,’ *kitto* ‘surely,’ *kippu* ‘ticket’, and that of /u/ in *suppai* ‘sour’ and *huttei* ‘scarcity’. Han (1962) notes devoicing of *kittari* ‘cutting’ and *suppai* ‘sour’. Nihon Onsei Gakkai (1976) includes examples such as *syuppan* ‘publication’, *hittyuu* ‘hitting the target’. According to Kawakami (1977), devoicing of /i/ is in principal obligatory in *kissaten* ‘coffee shop,’ *sippai* ‘failure’, and /u/ in *syuttoo* ‘appearance’, although

/u/ in *kyutto* ‘tightly’ can be voiced. The latter voiced case may be attributed to the accent on /u/.

However, quantitative studies showed that devoicing before geminates is not systematic but varies interpersonally (Han 1994; Kondo 2001; Maekawa and Kikuchi 2005; Shrosbree 2013). In the CSJ analysis (Maekawa and Kikuchi 2005), the devoicing tendency due to the manner of consonants is generally similar before geminates and before singletons. That is, the devoicing rate was highest for the combination of fricative and stop geminate (Fr-QSt), and lowest for that of fricative and fricative geminate (Fr-QFr). On the other hand, geminates and singletons exhibit a difference in the devoicing of their preceding vowel if they are preceded by stops: the devoicing rate is lower if stops are followed by geminate consonants (St-QC$_2$) than if they are followed by single consonants (St-C$_2$), regardless of the manner of C$_2$.

The glottal opening patterns of the two subjects in Sawashima’s (1969) experiment show mostly plateau-type or bimodal during devoiced /kitt/, /sitt/, /kiss/, and /siss/, whereas those that are observed during devoiced /kit/, /sit/, /kis/, and /sis/ are mono-modal. This suggests that the glottal manifestation differs between geminates and singletons when they are placed in the C$_2$ position. Considering the less frequent devoicing rate and the tendency towards bimodal glottal pattern, C-QC sequences can be categorized as atypical consonantal conditions. Hattori (1984) claims that glottal tension appears in the first half of geminate consonants. If so, this might lead the glottal opening pattern to be bimodal during C-QC. Physiological data which are currently available indirectly support this claim for stops and affricates, but not for fricatives (Fujimoto, Maekawa, and Funatsu 2010; Fujimoto, Funatsu, and Fujimoto 2011). Detailed examinations are necessary in order to evaluate Hattori’s (1984) claim as well as its relation to the devoicing of the preceding vowels.

2.2.4 Devoicing involving voiced segments

Vowels can be devoiced when they are followed by voiced segments. Han (1962) reports that high vowels are occasionally devoiced when they are preceded by voiceless consonants, [s] or [ʃ] particularly, and followed by a semivowel [j], as in *soo desuyo* ‘that’s right’. In the speech by a female announcer, high-vowels are devoiced before nasals as in [desɯ̥ŋa] ‘but (polite form)’ and [sɯsɯ̥mete] ‘to advance’, and even before a vowel as in [tsɯ̥ite] ‘as to’ (Maekawa 1983). Also in the speech data of a male announcer, devoicing was frequent when high vowels are followed by /g/ (Takeda and Kuwabara 1987). In a study which involved many speakers, devoicing before voiced consonants occurred in some speakers in the words [ʃi̥nabirɯ] ‘to wither’, [imaʃi̥merɯ] ‘to admonish’, [itaʃi̥masɯ] ‘to do (polite form)’, [itadaki̥masɯ] ‘thanks’, [sɯ̥masɯ] ‘to finish’, [sɯ̥nawatʃi] ‘that is’, [desɯ̥ga] ‘but (polite form)’, [desɯ̥ne] ‘isn’t it?’, and [masɯ̥node] ‘because (polite form)’, although such cases

are exceptional (Kawai et al. 1995). Many instances noted above relate to the words *desu* ‘copula (polite form)’ and *masu* ‘auxiliary verb (polite form)’, or when the following consonants are nasals. For the former cases, it is likely that the higher frequency of the words leads to the higher devoicing tendency.

Similar tendencies are reported in the corpus analyses. In the CSJ corpus (Maekawa and Kikuchi 2005), high vowel devoicing before voiced consonants occurred at 17.37% in /i/ and 20.91% in /u/. The devoicing rate is highest when the vowels are followed by nasals (35.8%) and second highest when followed by approximants (18.4%). Maekawa and Kikuchi note that the frequent occurrence of *desuyo* ‘*desu* + *yo*’, *masuyo* ‘*masu* + *yo*’, and *masuwa* ‘*masu* + *wa*’ idiosyncratically pushed up the average rate. In the corpus of telephone speech, devoicing of the first /u/ in *kuru* ‘to come’ and *suru* ‘to do’ is also reported (Komatsu and Aoyagi 2005; Arai, Warner, and Greenberg 2007).

In the CSJ corpus (Maekawa and Kikuchi 2005), devoicing of non-high vowels between voiceless C_1 and voiced C_2 occurred at the rate of 0.49%, 1.05%, and 1.81% for /a/, /e/, and /o/, respectively. Also, devoicing between voiced C_1 and voiceless C_2, as well as between voiced consonants, occurred for all vowels, although the devoicing rates are negligibly low (2.23% or less). In the telephone corpus, devoicing between voiced consonants such as /i/ in *hazime* ‘beginning’ is reported (Komatsu and Aoyagi 2005; Arai, Warner, and Greenberg 2007). However, these rare cases are found only in corpus data and may deserve a detailed examination. In the CSJ corpus, there are occasional instances where utterance-final phrases become totally voiceless. In the telephone corpus, speech signals lower than 300Hz were omitted due to the limitation of cut-off frequencies, which may result in losing very short and weak glottal pulses.

2.3 Summary of the segmental conditions

Based on the above studies, devoicing probability can be divided into several categories depending on the segmental conditions as shown in Table 2 below. Affricates are categorized together with fricatives in the C_1 position and with stops in the C_2 position. Devoicing in word/phrase-final conditions is separately discussed in section 2.7.1. The terms “general” and “non-general environment”, as well as “typical” and “atypical consonantal conditions” are used in the rest of this chapter. Note that these terms may be used differently in other studies. Especially, the term “typical condition” often denotes in other studies the environments where high vowels are placed between voiceless consonants, which is referred to by the term “general environment” in this chapter.

Table 2: Devoicing environments categorized by devoicing probability.

devoicing conditions		segmental sequences	examples	devoicing frequency in unaccented mora
general devoicing environments	typical consonantal conditions	St-St/Af, Af/Fr-St/Af, St-Fr (non-/h/)	kutu 'shoes' sika 'deer' kusa 'grass'	systematic/ highly frequent
	atypical consonantal conditions	Af/Fr-Fr, St/Af/Fr-/h/ C̥-V_h-QC̥,	susi 'sushi' sihei 'bill' sikki 'lacquer'	non-systematic/ moderately frequent
non-general devoicing environments		C̥-V_{nh}-C̥, C̥-V_h-C̮,	haha 'moher' desune 'isn't it'	non-systematic/ less frequent

2.4 Accent

The literature often states that accented vowels do not devoice (Kawakami 1977; Shibatani 1990; Hibiya 1999); see Kawahara (this volume) for details of Japanese word accent. Some acoustic studies indeed report that devoicing for accented vowels does not occur (Takeda and Kuwabara 1987) or rarely occurs (Kimura, Kaiki, and Kitoh 1998). However, many studies have shown that devoicing occurs moderately in accented vowels, although the frequency is lower than in unaccented ones (Han 1962; Fujisaki et al. 1984; Sakurai 1985; Kuriyagawa and Sawashima 1986; Yoshida and Sagisaka 1990; Sugito 1997; Kimura, Kaiki, and Kitoh 1998; Imai 2004; Fujimoto 2004a). Nagano-Madsen (1994b) found that accent as well as tone did not suppress devoicing in sentences, as compared to isolated words. Kondo (1997) reports that word accent does not suppress devoicing unless it is in consecutive devoicing environments. In Fujimoto (2004a), who looked at ten young male speakers, three speakers show more frequent devoicing in accented vowels than unaccented ones, but three speakers show similar rates, while the average devoicing rate is lower in the accented vowels. Thus, it appears that there is much inter-speaker variation in the realization of devoicing in accented vowels.

In an electromyographic study of a Tokyo speaker, excitations of PCA and INT are both greater in accented /si'hee/ than in unaccented /sihee/ (Yoshioka 1981). Since PCA is the (only) glottal abductor, greater excitation of the muscle may lead to larger glottal opening, which plausibly facilitates devoicing. On the other hand, greater excitation of INT, a glottal adductor muscle, leads to glottal closing, which suppresses devoicing. This suggests that, in accented syllables, both devoicing and voicing gestures are augmented. Whether or not devoicing actually occurs may depend on which muscle, the PCA or the INT, has stronger excitation. This may cause the variation in devoicing from token to token, or from person to person.

Han (1962) observed that devoicing is frequent on the mora adjacent to the accented one, either immediately preceding or following it. Between the two adjacent

positions, devoicing occurs more frequently at the moras that follow the accented mora than those that precede it (Sakurai 1985; Takeda and Kuwabara 1987; Kimura, Kaiki, and Kitoh 1998). These results suggest the likelihood of devoicing in low-pitched moras, since only low pitch is assigned after the accented mora while both high and low pitches are possible at the mora before it. This analysis is supported by Kuriyagawa and Sawashima (1986), Yoshida and Sagisaka (1990), and Imai (2004), who all demonstrated that low-pitched vowels are more likely to devoice than high-pitched ones.

2.5 Speech rate

2.5.1 General description and acoustic studies

Unlike English schwa deletion which occurs more frequently in faster speech (e.g. Dalby 1986), Japanese devoicing occurs regularly at a normal speech rate especially in typical consonantal conditions. However, in non-general devoicing environments, devoicing occurs in non-high vowels in fast speech as in the first /o/ in *kokoro* 'heart' (section 2.1.2). In acoustic studies, high vowels before voiced consonants (Maekawa 1989) and non-high vowels before voiceless and voiced consonants devoice in fast speech, but not in normal tempo (Maekawa 1990). In the CSJ corpus, non-high vowel devoicing is more frequent in faster speech, except for /o/ (Maekawa and Kikuchi 2005). Devoicing in consecutive devoicing environments also increases in fast speech, while that in single devoicing environments regularly occurs irrespective of speech rate (Kondo 1997).

In slower, elaborated speech, devoicing is suppressed (Han 1962, Imaizumi, Hayashi, and Deguchi 1995). Nevertheless, devoicing in typical consonantal conditions is not totally omitted even in the teachers' speeches that address hearing impaired children (Imaizumi, Hayashi, and Deguchi 1995). This implies the robustness of articulatory tendencies toward devoicing in typical consonantal conditions among Tokyo speakers.

2.5.2 Physiological studies

There are not many physiological studies on the effect of speech rate on devoicing. This is largely because devoicing regularly occurs at a normal speech rate especially in a typical consonantal condition (section 2.2.1), where the glottal opening patterns are mono-modal (section 2.2.2). The effect of speech rate can be seen in atypical consonantal conditions in which two glottal openings may concatenate or merge in fast speech. Figure 4 shows the glottal opening pattern of /kihe/ in normal and fast speech produced by a Tokyo speaker who is different from the speaker in Figures 1,

2, and 3. Note that /kihe/ has an atypical consonantal condition in which devoicing is non-systematic (section 2.2.1). This speaker shows no devoicing of /i/ for six repetitions at a normal rate and only once in six repetitions at a fast rate. At the normal rate (left), the glottal opening pattern (bottom waveform) shows two separate openings, each corresponding to /k/ and to /h/. The speech wave (top) shows continuous voicing during /ihe/. In the devoiced token at the fast rate (right), in contrast, the openings for /k/ and /h/ concatenate with a dip in between and the speech wave shows no voicing during /ih/. It clearly shows that this devoicing is due to the concatenation of two glottal openings, not to the glottal reorganization as was seen in Figure 1.

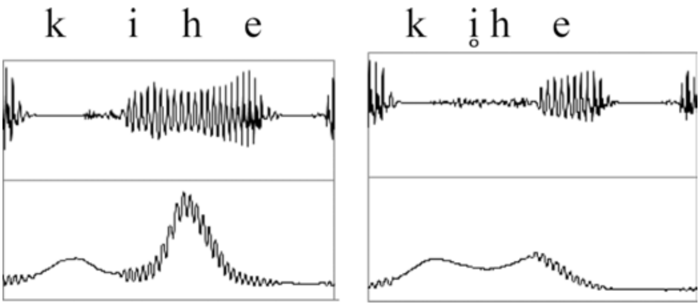


Figure 4: Speech wave (top) and glottographic signal (bottom) during the unaccented non-word /kihe/ in normal speaking rate (left) and fast speaking rate (right) produced by a Tokyo speaker. Vowel /i/ is devoiced in fast speech. (Adapted from Fujimoto, Funatsu, and Fujimoto 2012)

Munhall and Löfqvist (1992) revealed that the glottal openings of the /st/ sequence in English 'Kiss Ted' show two separate openings, each corresponding to /s#/ and /#t/, at slow speed, but the openings concatenate to be a bimodal pattern as the speech rate increases, and they merge into one in fast speech. Although intervened by a vowel, the glottal opening of two voiceless consonants in Japanese /CVC/ may behave in a similar manner. That is, the distance between the two independent openings of /C_1/ and /C_2/ in /C_1VC_2/ shortens as the speech rate increases. In faster speech, the two openings concatenate to show a bimodal pattern, and further merge into one showing a mono-modal pattern. The intermediate vowel may be devoiced when the glottal aperture is wide enough to suppress the initiation of vocal fold vibration.

2.6 Dialects

2.6.1 General description and acoustic studies

Frequency of devoicing often cues which dialect is being spoken (Kindaichi 1954). The literature agrees that devoicing is frequent in Tokyo and its surrounding Kanto

area, whereas it is less frequent in Kyoto-Osaka and the surrounding Kinki area (Sakuma 1929; Sibata 1988; Matsumori et al. 2012). Sakuma (1929: 231) notes that the devoicing in conventional environments as in /u/ in *kusa* 'grass' and /i/ in *tikara* 'power' is robust in Tokyo Japanese, adding that "if voiced, it will sound like Kyoto-Osaka or Kochi dialects." Some literature reports that devoicing is infrequent in western Japan (e.g. Mase 1977). However, the Western dialects such as Kagoshima in Kyushu as well as Okinawa further in the south show frequent devoicing (Hirayama 1985). In Sugito's (1996) investigation, the devoicing rate among seven cities is, from east to west, 55.6% in Sendai, 55.6% in Tokyo, 67.7% in Nagoya, 32.3% in Osaka, 28.0% in Okayama, 18.0% in Kochi (Shikoku), 53.2% in Kumamoto (Kyushu), and 56.7% in Naha, Okinawa. In this study, Kumamoto, Naha, and Sendai show devoicing rates similar to Tokyo. On the other hand, Nagoya, which is generally categorized as a less frequent devoicing area, shows the highest devoicing rate.

One must note that the accentual pattern differs considerably depending on the dialects, and this may strongly affect the devoicing rate. For example, many two-mora words with a H(igh)-L(low) accent pattern in the Kinki dialects have a L(ow)-H(igh) pattern in Tokyo Japanese (e.g. *kusa* 'grass' and *sika* 'deer'). Given that devoicing is less frequent on accented vowels (section 2.4), it naturally follows that devoicing of this word group becomes lower in Kinki dialects than in the Tokyo dialect, even if the potential devoicing behavior were the same.

Table 3 shows the devoicing rate of ten Osaka speakers when they produced unaccented non-words (Fujimoto 2004a). Table 3-1 shows the average of eight speakers who demonstrate frequent devoicing, while Table 3-2 shows that of two speakers who exhibit infrequent devoicing. These tables are comparable to Table 1 (section 2.2.1) which shows the devoicing rate of Tokyo dialect speakers. The devoicing rate of the Osaka speakers is, on average, lower than that of Tokyo speakers as pointed out by many previous studies. However, in Table 3-1, devoicing and its pattern due to consonantal conditions are similar to those in the Tokyo speakers. Namely, the devoicing of /i/ in unaccented /C_1iC_2e/ words is, consistent (98%) in typical consonantal conditions, and less frequent (33%) when the combination is 'Fr-Fr' (/sise/, /sihe/, /hise/, and /hihe/), and very low (10%) when C_2 is a /h/ regardless of the C_1 (/kihe/, /sihe/, and /hihe/). In Table 3-2, devoicing occurred only in typical consonantal conditions (41%), and none in others. These results further support the categorization of the general devoicing environments into typical and atypical consonantal conditions. As can be seen from these tables, inter-speaker variation is large among Kinki speakers, while Tokyo speakers demonstrate homogeneity especially in the typical consonantal conditions (Fujimoto and Kiritani 2003; Fujimoto 2004a).

Table 3: Devoicing rate for Osaka speakers of /i/ in unaccented /C_1iC_2e/ non-words uttered in a frame sentence. The top table (3-1) shows the average of eight Osaka speakers with frequent devoicing, and the bottom (3-2), two Osaka speakers with less frequent devoicing. (Adapted from Fujimoto 2004a)

3-1

C_1 \ C_2	k	t	s	h
k	94%	94%	97%	13%
s	100%	100%	53%	13%
h	100%	100%	63%	3%

3-2

C_1 \ C_2	k	t	s	h
k	13%	25%	38%	0%
s	75%	75%	0%	0%
h	0%	63%	0%	0%

It is worth noting here that the pronunciation of /u/ is claimed to differ among dialects. It is realized as unrounded [ɯ] in Tokyo and rounded [u] in Kinki (Umegaki 1968; Okumura 1975; Yamamoto 1982). If so, the duration of /u/ is plausibly longer in Kinki dialects due to lip rounding and/or lip protrusion in [u]. Then, the devoicing rate for /u/ may decrease in these dialects (Sugito 1996). So far, this claim has not been supported by acoustic or articulatory studies. Further examination is desirable on the phonetic variation of /u/ among dialects.

The lower devoicing rates among Kinki speakers may be attributed to their slower speech rate. However, this is not always true, since some speakers of Kinki dialects with a slower speech rate showed a similar devoicing frequency as Tokyo speakers (Sugito 1996; Fujimoto and Kiritani 2003). Further empirical studies are necessary on this issue, too.

In the Kagoshima dialect, word-final vowels are often lost. For example, *kaki* 'persimmon' is pronounced as [kat] and *kami* 'paper' as [kaN] (Sibata 1988).[1] This may look similar to word-final devoicing of vowels, but it differs from devoicing in that it deletes the word-final vowels completely rather than devoices them (Sibata 1988).

2.6.2 Physiological studies

Among Tokyo speakers, only mono-modal glottal opening patterns appear for devoicing in typical consonantal conditions (section 2.2.2). In other dialects, however, bimodal patterns appear in devoiced /CVC/ even in typical consonantal conditions

1 It is an interesting question why they don't say [kak] and [kam].

(Fujimoto 2005; Fujimoto et. al. 2010; Fujimoto, Funatsu, and Fujimoto 2012; Fujimoto 2012). Figure 5 shows the glottal opening pattern during /kise/ (top) and /sike/ (bottom) as produced by an Osaka dialect speaker. The vowel /i/ is devoiced in all tokens except for the first (top, leftmost) /kise/. Thus, the devoicing rate of this speaker is very high: 75% for /kise/ and 100% for /sike/, an average of 88% across both words. However, the glottal opening pattern for the devoiced tokens is bimodal (second and third /kise/ and second and third /sike/) or plateau-type (fourth /kise/ and the first and fourth /sike/). These bimodal patters can be interpreted as the concatenated shape of the C_1 and C_2 openings. Namely, these glottal opening patterns are not reorganized. Also, the degree of the glottal opening in the devoiced tokens is often comparable to that of single consonants in the voiced tokens.

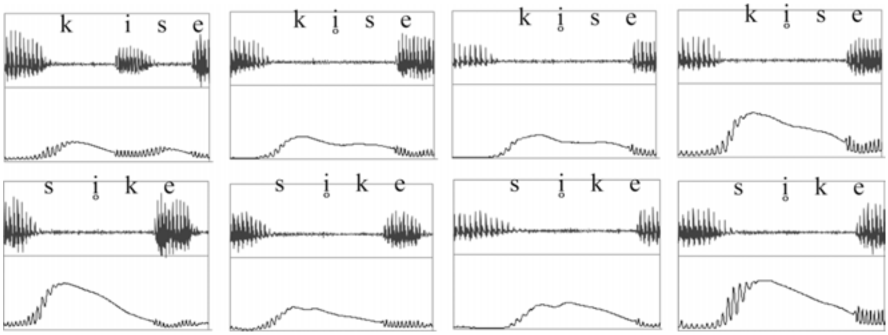


Figure 5: Speech wave (top) and glottographic signal (bottom) during the unaccented non-words /kise/ (top) and /sike/ (bottom) produced by an Osaka speaker. Vowel /i/ is devoiced in all tokens except for the first /kise/ token (upper left). (Modified from Fujimoto 2005)

Electromyographic data of another Osaka dialect speaker give additional evidence for this observation. Although this speaker shows a high devoicing rate (around 90%), the glottal opening muscle (PCA) showed two separate activations, each corresponding to the preceding and the following consonants (Fujimoto et al. 2005; Fujimoto 2006).

2.7 Position in words

2.7.1 Phrase-final devoicing: general description and acoustic studies

Another well-known condition of devoicing concerns high vowels between voiceless consonants and a pause. According to Nihon Onsei Gakkai (1976), devoicing always occurs in this environment. Sakurai (1985) notes that word-final high vowels devoice when they are in low-pitched unaccented moras. In contrast, Sibata (1988) states that high-vowel devoicing in word-final position is not common in the Tokyo dialect, although it is frequent in the Kumamoto dialect in Kyushu and some others (e.g. *kaki* 'persimmon' [kaki̥] and *kami* 'paper' [kami̥]).

The results of acoustic analysis differ from study to study. In Han (1962), phrase-final devoicing is salient in the words *desu* 'copula (polite form)' and *masu* 'auxiliary verb (polite form)', which are used very frequently in everyday conversation. Similarly, phrase-final /u/ in *hai soodesu* 'yes, that's right' almost always devoices in the Tokyo dialect (Maekawa 1989). Also, pre-pausal low-pitched vowels are uniformly devoiced by younger speakers (Imai 2004). In Fujimoto and Kiritani (2003), four out of five Tokyo speakers showed consistent devoicing of /u/ in sentence-final *masu*. The other speaker showed consistent devoicing in one carrier sentence and consistent weak voicing in the other.

On the other hand, in isolated words read by a male announcer, low-pitched word-final high vowels are devoiced only in one out of 29 instances (Takeda and Kuwabara 1987). This is contrary to Sakurai's (1985) description and Imai's (2004) result. In Kimura, Kaiki, and Kitoh (1998), too, word-final devoicing rarely occurred. Also, in Kawai et al. (1995), devoicing before a pause is not consistent, although ten speakers show devoicing to some extent. From their data, the average rate of word/phrase-final devoicing is 11%, which is far from being a regular occurrence (Kawai et al. 1995). In Byun (2007), too, devoicing of phrase-final /u/ is not systematic in many dialects including Tokyo.

As was shown in many studies, devoicing between a voiceless consonant and a pause does not consistently occur. Interpersonal variation may be one cause. More importantly, however, the above results suggest that devoicing is more likely to occur phrase-finally than word-finally. In phrase-final position, function words such as *desu* and *masu* often appear, in which devoicing is very common. In word-final position, in contrast, word frequency effect is not much expected. In addition, speakers may pronounce words elaborately when uttered in isolation, which may decrease the devoicing rate word-finally.

Pitch of the mora is another factor that affects devoicing. Nihon Onsei Gakkai (1976) remarks that devoicing does not occur when the pitch of the last mora is higher than that of the preceding one (e.g. /su/ in *bo'ku wa osū* 'I push'). In acoustic analysis, devoicing is generally less frequent when the pitch of the relevant mora is high, as compared to when it is low (Kuriyagawa and Sawashima 1986; Yoshida and Sagisaka 1990; Imai 2004). However, in Fujimoto and Kiritani (2003) cited above, /u/ of sentence-final *masu* by a Tokyo speaker showed consistent devoicing in one carrier sentence and consistent voicing in the other, although the target moras *su* were both low-pitched. The pitch contour of the carrier sentences in the Tokyo dialect are LHH HLL LH<u>HL</u> for the devoiced case, and LHH HHH HH<u>HL</u> for the voiced case (underline denotes the pitch of the sentence-final *masu*). Namely, the voiced case is preceded by a longer sequence of high-pitched moras than the devoiced case. This may be interpreted as suggesting that longer sequences of high-pitched moras suppress devoicing of the immediate following mora. Detailed investigation of the effect of pitch contours on devoicing is desirable. It is worth mentioning that, in the voiced case, the vowel showed a shorter duration and

weaker intensity than full vowels (Fujimoto and Kiritani 2003). Namely, the vowel was reduced. Figure 6 shows the examples of the speech wave and the spectrogram of the voiced tokens as produced by a Tokyo and a Kinki speaker. The vowel /u/ by the Tokyo speaker is reduced, or partially devoiced. The reduced vowels will be discussed later in section 4.1.

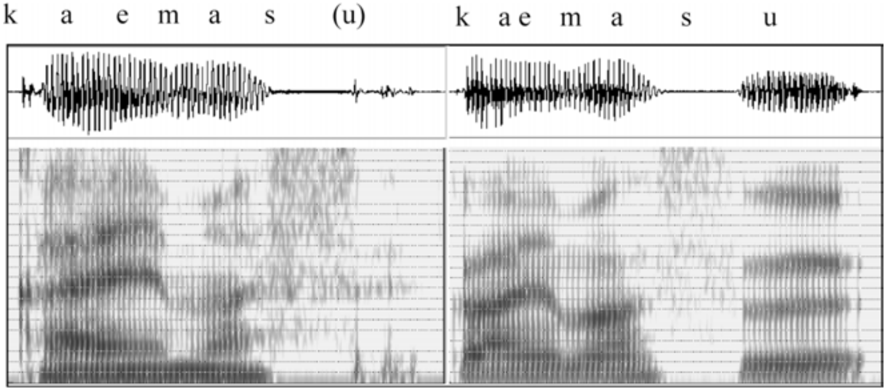


Figure 6: Speech wave and spectrogram of /kaemasu/ uttered by a Tokyo speaker (left) and by a Kinki speaker (right). The sentence-final /u/ by the Tokyo speaker (left) is reduced and that by the Kinki speaker (right) is fully voiced. (Adapted from Fujimoto and Kiritani 2003)

Intonation and boundary pitch movements are also very strong factors that can affect word/phrase-final devoicing. Devoicing does not occur when the last mora carries a rising intonation, as in interrogative sentences (Nihon Onsei Gakkai 1976; Kawakami 1977). Maekawa (1989) observed that, for the Tottori dialect in the Chūgoku region, devoicing was infrequent in the phrase-final /i/ as in /*…notoki*#/ 'when…' and /*ase mo kakanaisi*#/ 'do not even sweat'. He attributes this lower devoicing rate to the (non-lexical) pitch rise in these final moras. Hence, in addition to lexical accent, pitch rise due to sentence-level intonation such as interrogation and emphasis should be taken into consideration.

In sum, word/phrase-final devoicing is not as systematic as was traditionally described. Devoicing is plausibly more frequent phrase-finally than word-finally. The effects of word frequency, word type (i.e. content vs. function words), accentual patterns and intonation should all be taken into account. Further investigation is essential in order to clarify the details of word/phrase-final devoicing.

2.7.2 Phrase-final devoicing: physiological studies

There are not many physiological studies about word/phrase-final devoicing. Figure 7 gives examples of sentence-final *masu* produced by two Tokyo speakers and one

Osaka speaker (Fujimoto, Funatsu, and Fujimoto 2012). As can be seen in the speech signal, /u/ is fully devoiced for a Tokyo speaker (left), partially devoiced for another Tokyo speaker (middle), and fully voiced for the Osaka speaker (right). In the fully devoiced case (left), the glottis continuously opens from /s/ to the inhalation position, showing no closing movement for /u/. This glottal opening pattern is observed consistently in this speaker's utterances. This consistency suggests that the gesture for /CV#/ is phonologically reorganized into /C#/, namely /masu/ → /mas/.

On the other hand, the glottal opening pattern of the other Tokyo speaker (middle) shows a closing movement for /u/ after /s/. This pattern is observed consistently in this particular speaker, which suggests that the speaker intended to produce /u/ as a full vowel. In this case, the vowel is partially devoiced in the sense that it has reduced intensity and duration. This partial-devoicing may be achieved by factors other than glottal opening, such as tighter oral closure or glottal tension. These data suggest that both phonological and phonetic (categorical and non-categorical) types of devoicing are involved in phrase-final devoicing even among Tokyo speakers. This agrees with the acoustic findings that the extent of phrase-final devoicing varies from person to person and from utterance to utterance. Finally, the Osaka speaker, in the right, shows a glottal opening pattern similar to the Tokyo speaker in the middle, but /u/ is fully voiced in the speech signal. Further physiological studies are required to clarify what happens with word/phrase-final devoicing.

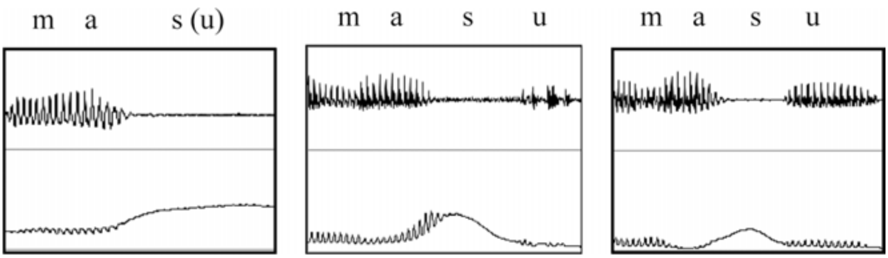


Figure 7: Speech wave (top) and glottographic signal (bottom) during phrase-final /masu/ produced by two Tokyo speakers (left and middle) and an Osaka speaker (right). The vowel /u/ is fully devoiced in the left, partially devoiced in the middle, and fully voiced in the right panel.

2.7.3 Phrase-initial devoicing

Devoicing occurs frequently at word/phrase-initial position. Many examples cited in the literature report devoicing in word-initial moras for both high vowels, as in *kusi* 'comb' and *kisi* 'shore' (Han 1962), and non-high vowels, as in *kakaru* 'hang on' and *kokoro* 'heart' (Sakuma 1929). Acoustic analysis showed that devoicing is indeed more frequent word-initially than word-medially (Fujisaki et al. 1984; Kimura, Kaiki, and Kitoh 1998).

As for devoicing in non-high vowels, Kawakami (1977) remarks that voiceless stops are aspirated in word-initial position in Japanese. Thus, *kome* ‘rice’ is actually [k^{h}ome], where the longer aspiration period of [k^{h}] than [k] shortens the duration of [o] by the period (Kawakami 1977). Physiological studies have shown that glottal opening for voiceless obstruents is larger word-initially than word-medially in Japanese (Sawashima 1971a; Sawashima and Miyazaki 1973; Sawashima and Niimi 1974). It is therefore reasonable to assume that the larger glottal opening invades the following vowels, thereby shortening their durations. This leads to the tendency for devoicing in word-initial moras compared to word-medial and final moras.

2.8 Consecutive devoicing

2.8.1 General description and acoustic studies

In cases where two or more devoiceable moras adjoin, devoicing may or may not occur in all of the target vowels. Sakuma (1929) notes that devoicing consecutively occurs in words such as *kutisaki* [kɯ̥tʃi̥saki] ‘lips’ and *sitisyaku* [ʃi̥tʃi̥ʃakɯ] ‘seven shaku (a unit of scaling distance)’ unless uttered in slow tempo. Nihon Onsei Gakkai (1976) notes that consecutive devoicing may be avoided in case too many devoiced vowels leads to confusion by the listener, as in [ki̥kɯtʃi̥kaɴ] or [ki̥kɯ̥tʃikaɴ] for *Kikuchi Kan* (novelist’s name), [reki̥ʃiteki] for *rekisiteki* ‘historical’, and [ɸɯ̥kɯʃi̥kikokjɯː] for *hukusikikokyuu* ‘abdominal breathing’.

McCawley (1968: 127) notes that “when several consecutive syllables each contain a diffuse short vowel between voiceless consonants, only alternate vowels become voiceless. However, whether the first, third, fifth, etc., or the second, fourth, etc. become voiceless depends on several factors such as which vowels are /i/’s and /u/’s and what the consonants are”. Sakurai (1985) remarks that devoicing does not occur in one of the two devoiceable mora as in *kikikata* [ki̥kikata] ‘listener’ and *takitukeru* [takitsɯ̥kerɯ] ‘to kindle’, and, in the middle mora of three devoiceable sequences as in *kikisuteru* [ki̥kisɯ̥terɯ] ‘to ignore’ and *kikitukeru* [ki̥kitɯ̥kerɯ] ‘to overhear’.

Kawai et al. (1995) surveyed NHK’s pronunciation dictionary (1985), in which devoiced moras are supposedly marked. They found that 63.5% of the voiceless moras in two consecutive devoiceable environments turned out to be a devoiced-voiced (D-V) sequence and 35.4%, a voiced-devoiced (V-D) sequence; 84.7% of three consecutive devoiceable environments is described as a devoiced-voiced-devoiced (D-V-D) sequence; four consecutive devoiceable environments all turn into devoiced-voiced-devoiced-voiced (D-V-D-V) sequences. Thus, two voiced or two devoiced sequences are very rare in any consecutive environment. These results suggest that devoicing is favored in the first and the third syllables.

On the other hand, the results of acoustic analyses are inconsistent. Han (1962) observed that devoicing occurred in the first and third vowels or the second and fourth vowels if the consonants of consecutive moras are all stops. Sugito (1996) and Imai (2004) found that vowels in the two consecutive devoicing environments are often consecutively voiced, whereas Varden (2010) reports that they are often consecutively devoiced.

An analysis of the CJS corpus provides a more accurate picture (Maekawa and Kikuchi 2005). This corpus contains 318 word-internal consecutive devoicing environments of which 84 were consecutively devoiced, 17 were consecutively voiced, and the remaining 215 showed devoicing on one of the vowels. Thus, devoicing on alternating moras seems most common. Moreover, among the 215 cases showing devoicing on only one vowel, 171 involve a devoiced-voiced (D-V) sequence, whereas 44 involve a voiced-devoiced (V-D) sequence. This means that if consecutive devoicing does not occur, first vowels are more likely to devoice.

The same analysis of the CJS corpus furthermore reveals that the manner of the consonants plays an important role if devoicing occurred in only one vowel in the consecutive devoicing environments. If a fricative was combined with an affricate or stop, the vowel after the fricative was more often devoiced. If both consonants are fricatives, the vowel after the second fricative was more often devoiced. Although Maekawa and Kikuchi (2005) do not mention, the following consonants would play an important role in this second case: i.e. C_3 in /$C_1VC_2VC_3$/. Suppose that C_3 is a stop or an affricate, it forms a Fr-Fr-St/Af sequence as in *sisitoo* [ʃiʃitoo] 'green pepper'. This condition has an atypical consonantal condition for the first vowel and a typical one for the second vowel. In such a case, it is likely for the vowel in the typical consonantal conditions to devoice. If C_3 is a fricative, in contrast, it forms a Fr-Fr-Fr sequence as in *susuharai* 'sweeping the soot off'. In this case, it is more likely for the first vowel to devoice because, as noted above, the glottal opening is generally larger in the initial consonant than in the medial one. As for the avoidance of consecutive devoicing, Kondo (1997) argues that consecutive devoicing is disfavored because it would violate a constraint in Japanese syllable structure (see Kondo 2005 for detailed analysis). This phenomenon is analyzed by Tsuchida (2001) in the framework of Optimality Theory.

2.8.2 Physiological studies

While acoustic results demonstrated that consecutive devoicing is infrequent, physiological studies are scarce and inconclusive. Figure 8 gives a rare example in which devoiced and partially devoiced vowels are observed. It shows the glottal opening pattern of the unaccented non-word /kikiki/ produced by a Tokyo speaker. It is uttered in isolation in a normal speech rate for the purpose of high-speed digital recordings. Speech and PGG signals are overlaid in this figure so that the relation

of the two signals can be easily seen. In this particular token, the first /i/ is fully devoiced, while the second /i/ shows a couple of vibrations. Namely, the second /i/ is partially devoiced. The glottal opening pattern is mono-modal, similar to the reorganized pattern in Figure 1. More importantly, the range of the glottal opening here is somewhat longer than that of /C_1VC_2/, which eventually leads to the partial devoicing of the second vowel. However, the opening does not directly stretch into /C_3/. Namely, the scope of the glottal reorganization is /C_1VC_2/, not /$C_1VC_2VC_3$/. Thus, the mechanism of devoicing of the first and the second vowels differs. So far, the glottal reorganizing pattern for /$C_1VC_2VC_3$/ has not been found. Consecutive devoicing may be achieved when a reorganized glottal opening for /C_1VC_2/ is large enough to invade completely into the following vowel. If so, it is reasonable that the duration and the intensity of the second vowel differ from token to token depending on the duration of the glottal openings. This explains the non-systematic occurrence of consecutive devoicing.

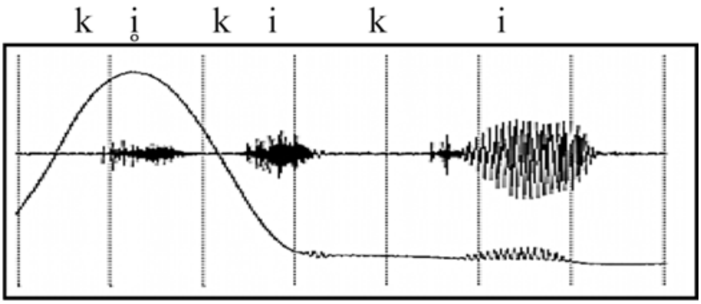


Figure 8: Glottal opening pattern during unaccented non-word /kikiki/ produced by a Tokyo speaker. Speech wave and glottographic signal are superimposed.

Figure 9 shows selected glottal images of a high-speed movie of the same /kikiki/ utterance in Figure 8. The sampling rate of the recordings is 4500 frames /second. In each frame, the vertical line on the speech wave denotes the timing of the excerpt images. Frames (a) to (d) correspond to the word-initial voiceless /kik/ sequence, frames (e)–(i) to the partially devoiced vowel of the second /i/, frames (j)–(m) to the voiceless part of the third /k/, and frames (n)–(r) to the voiced cycles of the word-final /i/. As is clear from (e) to (i), a couple of cycles of glottal vibration occurred which correspond to the speech and PGG signal in Figure 8. This kind of short voicing can easily be detected if the following consonant is a stop or an affricate, since they have silent part in the beginning. However, such short voicing can be overridden by the frication noise if the following consonant is a fricative. Further examination is essential in order to understand the mechanism of consecutive devoicing.

It is worth noting that the glottal opening for the third /k/ is negligibly small. In the frames (k) to (m) in Figure 9, the glottis stays very narrowly opened, at

the degree of half way between the open and closed phase of glottal vibration in (o) to (r). Narrow glottal opening in word-medial and word-final stop consonants is repeatedly observed regardless of subjects (Fujimoto 2004b). Nevertheless, the consonants are realized as voiceless, owing presumably to the supralaryngeal articulation. This may be a facilitative factor of stop consonants in the following position on the devoicing of the preceding vowel, which is contrastive to the voicing tendency of /h/ (section 2.2.2) (Fujimoto 2012).

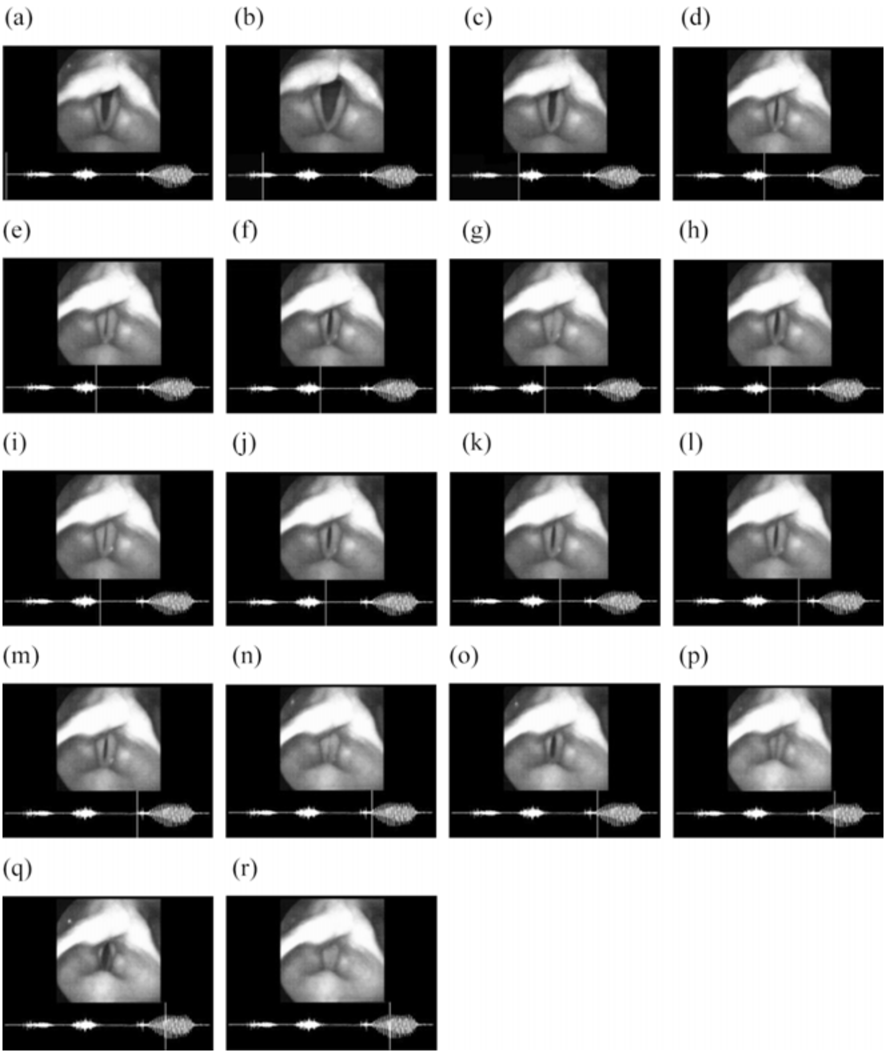


Figure 9: Glottal images excerpt from a high-speed movie of the /kikiki/ produced by a Tokyo speaker shown in Figure 8. Vertical line on the speech wave show the timing of the images in each frame.

2.9 Age, gender, social class and style

Non-linguistic or sociological factors also affect devoicing. Devoicing is infrequent among infants at the beginning of the language acquisition period. Interestingly, devoicing rates increase to an adult level around 4–5 years of age in the Tokyo dialect, whereas both infants and adults show infrequent devoicing in the Osaka dialect (Imaizumi et al. 1999).

Hirayama (1998) observes that devoicing is decreasing in the younger generation in Tokyo Japanese. However, in acoustic analyses, no generation gap was found among Tokyo speakers (Byun 2007, 2010). Imai (2010) investigated three age groups and found that young males devoiced most and young females, least, whereas the devoicing rate of middle and older age groups fell in between, showing little gender difference. Thus, the effect of generation is not straightforward and there may be some interaction with gender. When limited to the younger generation, males show more frequent devoicing than the females of the same age group (Varden 2010). This is consistent with Imai's (2010) findings. Akinaga (1985) argues that devoicing on accented vowels increases among younger generation. Nakao, Hibiya, and Hattori (1997) compared the old and new versions of NHK accent dictionaries and pointed out that the new version includes more devoiced vowels in accent moras.

Speech style and speakers' attitudes greatly affect devoicing, too. Devoicing is suppressed in elaborated, careful speech. Fujisaki et al. (1984) found that inter-speaker and intra-speaker variations depend on speakers' attitudes. In the task of reading 100 city names, four out of ten speakers show no devoicing at all, although one of them showed 70% devoicing when reading 1000 city names. This low rate of devoicing is due partly to the relatively slow (3 moras/sec) speech rate in all speakers, as well as to elaborated mora-by-mora pronunciation by a speaker (Fujisaki et al. 1984). Teachers use more voiced vowels when they address hearing impaired children as compared to normal hearing children (Imaizumi, Hayashi, and Deguchi 1995). However, mothers show virtually the same devoicing rates in infant-directed speech as those for adult-directed speech (Fais et al. 2010), which would help Tokyo children increase their devoicing rate to the adult-level by 4–5 years old, as Imaizumi et al. (1999) found.

Devoicing is highly frequent in spontaneous, conversational speech as compared to controlled speech (Komatsu and Aoyagi 2005; Arai, Warner, and Greenberg 2007; Imai 2010). Moreover, devoicing is more frequent in reading sentences than isolated words (Yasuda and Hayashi 2011; Shrosbree 2013). In sentences, word accent does not suppress devoicing according to Nagano-Madsen (1994a), although many studies claim it does (see section 2.4). The effect of speech style is more apparent in atypical consonantal conditions. In Shrosbree (2013), devoicing is less frequent in sentences than in isolated words only when the following consonants are geminates, whereas, in typical consonantal conditions, the rate is nearly 100% for both sentences and isolated words.

According to Maekawa and Kikuchi's (2005) analysis of the CSJ corpus, devoicing is more frequent in simulated public speech (with casual settings) than in academic presentation. Moreover, devoicing is more frequent when the vowels are uttered with laughter than without. Maekawa and Kikuchi claim that the presence of laughter indicates speaker's relaxation, resulting in a casual speaking style. There seems to be no physiological study that addresses this issue.

3 Perceptual studies

3.1 Perception of devoiced vowels

Given the facts about vowel devoicing in Japanese, one may wonder if listeners might have difficulties in identifying devoiced vowels as well as discriminating between different moras with the same onset consonants (e.g. /ki/ vs. /ku/). However, it is not the case for Japanese since the same consonantal phonemes differ phonetically before /i/ and /u/ due to co-articulation. Such differences are often specified in phonetic alphabets such as [tʃi] vs. [tsɯ] for /t/ and [çi] vs. [ɸɯ] for /h/. The same phonetic symbol is generally used for /k/, but this consonant is realized as palatalized [kʲ] before /i/ and non-palatalized [k] before /u/ (Kawakami 1977; Vance 2008). Plausibly, the same is true for /p/. Consonants in *yōʼon*, or palatalized sounds such as /sju/ [ʃɯ], /tju/ [tçɯ], and /hju/ [çɯ] also differ in articulation from those of *chokuʼon*, or single kana sounds, such as /si/ [ʃi], /ti/ [tʃi], and /hi/ [çi] (Kawakami 1977; Maekawa 1989). Coarticulation between a consonant and a vowel can be found in other languages, too, but the effect is particularly strong in Japanese, and this makes it easier for listeners to identify moras with devoiced vowels (Maekawa 1989).

Acoustic examinations confirm that spectral characteristics differ by and large between devoiced [ʃi] and [ʃɯ], and that such cues are readily perceived by native Japanese listeners (Beckman and Shoji 1984; Tsuchida 1994; Faber and Vance 2001). The degree of perceptual recoverability shows gradation depending on the amount of vocalic spectral information in the preceding consonants (Beckman and Shoji 1984). These studies examined fricatives, which are known to have a longer turbulent noise, i.e. a more prominent acoustic cue to the following vowel, than stops and affricates. In contrast, acoustic cues to the following devoiced vowel in the stops and affricates are not well documented, except for Yoshida (2008), who suggests that the similar acoustic cues are present in the VOT in /p/.

In spite of the fact that devoicing does not adversely affect perception of moras, devoicing is avoided in some conditions. When vowels consecutively devoice over several moras, sonority diminishes and the words become obscure. In such cases, only one of the vowels is voiced (section 2.8.1). In a single devoicing environment, devoicing between fricatives is less frequent (section 2.2.1). Devoicing is also less frequent if the following consonant is a stop geminate rather than a singleton stop (section 2.2.3). Maekawa and Kikuchi (2005) and Maekawa (2011) argue that devoic-

ing in these environments is avoided because it results in a succession of frication noise in the former case, and a succession of silent periods in the latter case, causing perceptual difficulties in detecting mora boundaries embedded within a stretch of voiceless sounds. As noted in section 2.2.1 above, Kawatsu and Maekawa (2009) and Maekawa (2011) showed that devoicing is less frequent if the cepstrum distance of the neighboring consonants is small, namely, if the two consonants are sound more similar, as in the Fr-Fr case. These results suggest that the speakers are monitoring their speech and controlling the voicing of the vowels. Actually, many Japanese are able to discriminate words with voiced and devoiced vowels (Funatsu et al. 2011), contrary to what Dupoux et al. (1999) reported. Also, since dialects vary according to their devoicing rates (section 2.6), judgments about dialects are by and large affected by the voicing of vowels (Morris 2010).

3.2 Detecting accent location on devoiced moras

Devoicing can occur in accented vowels (section 2.4). It is generally assumed that detecting an accent (or accented mora) becomes difficult if the accented vowel is devoiced. Traditional studies claim that listeners perceive accent on devoiced moras on the basis of the pitch fall of the vowel in the following mora (Hattori 1928; Kawakami 1969). Sugito (1982, 1997, 1998, 2003) argues that in the word *ku'sa* 'grass' in the Osaka dialect, the F0 falls sharply during /a/ regardless of the voicing status of /u/. Hasegawa and Hata (1992) confirmed this claim. Sugito and Hirose (1988) found that electromyographic patterns of Osaka speakers are similar in /ku'sa/ with or without devoicing of the accented /u/. That is, the neural command in realizing the accent is the same whether or not the target vowel is devoiced. Based on these findings, Sugito (1998, 2003) asserted that the sharp pitch fall in the following mora cues the presence of accent on the preceding devoiced vowel. It must be noted, however, that the inventory of pitch patterns is more complicated in the Kyoto-Osaka dialects than in the Tokyo dialect. Hence, it is not clear if Sugito and Hirose's (1988) electromyographic result can be generalized across dialects. Kitahara and Amano (2001) argue against the traditional view that accent location cannot be detected from the pitch contour. In general, pitch information can be perceived even in whispered speech, namely, in speech lacking an F0 contour. This issue deserves further investigation with regard to which acoustic information cues in perceiving accent on devoiced moras.

4 Other Issues

4.1 Gradient nature of devoicing

Voiced vowels in devoiceable environments are often short in duration and weak in intensity as compared to full vowels (Maekawa 1990; Kondo 1997, 2005). Such reduced

vowels are not fully voiced nor fully devoiced. They are referred to partially devoiced or half devoiced vowels. The duration and the intensity of partially devoiced vowels differ from token to token according to Kondo (1997), who includes such vowels in the words *desaki̱kikan* [desaki̱ki̥kan] ‘district office’ and *kutu̱*(*koozyoo*) [kɯ̥tsɯ̱] ‘shoe (factory)’ (partially devoiced vowels are denoted by an underline). Note that these vowels occur in consecutive devoicing environments.

Recall that partial devoicing of the second /i/ in /kikiki/ [ki̥kiki] in Figure 8 above is also observed in the consecutive devoicing environment, occurring at the end of a mono-modal glottal opening (section 2.8.2). Recall also that the voiced /u/ for sentence-final *masu* produced by a Tokyo speaker is partially devoiced (Figure 6 in section 2.7.1), but the glottal opening pattern for *masu* is not reorganized (Figure 7 in section 2.7.2). Empirical data on partially devoiced vowels are still sparse, but it may be possible to infer that such vowels are more likely to occur in environments other than typical consonantal conditions. Moreover, at least as far as currently available data for the Tokyo speakers are concerned, partially devoiced vowels are often produced with closed or narrowly opened glottis as shown in Figures 8 and 9.

4.2 Mechanism of devoicing

Mechanism of devoicing has been a long-standing issue. During speech production, devoiced vowels may occur at two different levels: at the speech planning level, and at the process of articulatory execution. In the former, a voiced or devoiced vowel is selected and is executed as it is by motor commands. In this case, devoicing is intentional and categorical, and the produced vowels would either be fully voiced or fully devoiced. In the latter case, the voiced vowel is uniquely selected at the speech planning level. But the vowel loses voicing during speech production due to the assimilation of the neighboring consonants. In this case, devoicing is unintentional and non-categorical. If the vowel is devoiced in this unintentional manner, the acoustic output of the vowel in the speech signal may vary from utterance to utterance.

As was seen in the previous sections, Tokyo speakers’ devoicing in /C̥V_hC̥/ environments is twofold: categorical devoicing as exemplified by typical consonantal conditions, and non-categorical devoicing as exemplified by atypical consonantal conditions (section 2.2.1). In typical consonantal conditions, the glottal opening is reorganized into a mono-modal pattern. In atypical consonantal conditions, the glottal openings often show bimodal or plateau patterns (section 2.2.2). However, when other dialects are included, bimodal glottal openings ubiquitously appear regardless of the consonantal conditions (section 2.6.2). Hence, Japanese devoicing as a whole can be basically explained by the assimilation of glottal gestures of voiceless segments onto a vowel. This is consistent with Sakuma’s (1929) description

of Japanese vowel devoicing as being executed by voicing assimilation where vowels lose voicing by the effect of the neighboring voiceless consonant(s).

The question remains how the reorganized mono-modal pattern uniquely corresponds to typical consonantal conditions in the Tokyo standard Japanese. As mentioned in section 2.5.2, the independent glottal openings of /s#/ and /#t/ in *Kiss Ted* concatenate to show a bimodal opening pattern as the speech rate increases, and finally merge into mono-modal in fast speech (Munhall and Löfqvist 1992). A similar process can supposedly occur in Japanese /CVC/, if the glottal opening of Cs is large and closely placed. Cross-linguistic studies revealed that consonants which require aspiration or frication noise tend to have glottal openings of their own (Löfqvist and Yoshioka 1980, 1981; Yoshioka, Löfqvist, and Hirose 1981). If we assume that frication and aspiration of voiceless consonants are stronger in Tokyo than in some other dialects, as is often claimed in the literature (section 4.7), the glottal openings are plausibly larger in the Tokyo dialect. Then, it is conceivable that the preceding and the following glottal openings would merge into mono-modal more readily in Tokyo than in other dialects. The results of acoustic and physiological studies of Osaka dialect seem to support this analysis (section 4.7). Empirical physiological studies will be essential for the evaluation of these assumptions.

4.3 Phonological or phonetic

Scholars often argue whether devoicing is phonological or phonetic. In such a case, phonological devoicing refers to the planned, categorical choice, whereas phonetic devoicing refers to a unintentional, random occurrence. Browman and Goldstein (1990, 1992) demonstrated that phenomena which are regarded as phonological events can be explained as purely phonetic ones. They showed that t-deletion in the phrase *perfect memory* occurred in speech signal, although the tongue gesture for /t/ is actually present. Hence, /t/ is not deleted but overlapped and hidden by the lip gesture. Based on Browman and Goldstein's (1990, 1992) gestural overlap analysis, Jun and Beckman (1993) and Beckman (1996) advocated that Japanese devoicing is purely a phonetic phenomenon which occurs due to glottal gesture overlap. In their assumption, the glottal opening gesture of the preceding (and the following) voiceless consonant(s) overlaps with the glottal closing gesture of the vowel, which reduces the duration and intensity of the vowel. The degree of reduction varies depending on the size and the timing of the overlapping glottal gesture onto the vowel. The random occurrence of devoicing in atypical consonantal conditions and bimodal glottal opening in these environments suggest that devoicing occurs due to this unintentional manner. This analysis is compatible with the occurrence of partially devoiced vowels.

On the other hand, systematic occurrences of devoicing and mono-modal glottal opening for devoiced /CVC/ along with the single excitation of the glottal opening muscle in typical consonantal conditions observed among Tokyo speakers are robust counter examples against the gestural overlap analysis. Currently, it seems safe to postulate that Japanese devoicing involves both phonological and phonetic aspects, as far as Tokyo Japanese is concerned. Tsuchida (1997), based on her physiological study, also concludes that Japanese devoicing has both phonological and phonetic aspects. In contrast, Kondo (1997) argues that neither traditional phonological accounts of devoicing based on the categorical process, nor phonetic accounts based on the gestural overlap analysis satisfactory explain the whole process of vowel devoicing in Japanese.

4.4 Devoiced or deleted

Some scholars distinguish between devoiced and deleted vowels to explain intra-personal variations. Sakuma (1929) claims that at the phrase/word-final position, vowel articulation is totally omitted leaving only "bare" consonants in words like *soodesu* 'that's right' and *kasi* 'confectionery', whereas it is kept in words like *tenpi* [tempi̥] 'oven'. Jinbō and Tsunemi (1932) note that if the first vowels are deleted, *susumu* 'to go forward' and *sisya* 'messenger' will become [ssmu] and [ʃʃa].[2] According to Kawakami (1977), vowels are devoiced in /ki/ [ki], /pi/ [pi], /ku/ [kɯ], /pu/ [pɯ], /sju/ [ʃɯ], /tju/ [tʃɯ] moras; they are deleted or, extremely short, in /si/ [ʃi], /ti/ [tʃi], /hi/ [çi], /su/ [sɯ], /tu/ [tsɯ], /hu/ [ɸɯ] moras; but they are not necessarily devoiced in /pju/ [pjɯ], /kju/ [kjɯ], /hju/ [çjɯ] moras. He claims that *sikaku* 'qualification' is usually pronounced as [ʃkakɯ], whereas *syukaku* 'nominative case' is always [ʃɯ̥kak]. He also argues that vowel devoicing is preferred over vowel deletion if the same consonant precedes and follows a vowel as in *sisyoku* [ʃi̥ʃokɯ] 'tasting' and *sissyoku* [ʃi̥ʃʃokɯ] 'unemployment'. Maekawa (1989) notes that /i/ is devoiced in *akikara* 'from autumn', but deleted in *asita* 'tomorrow'. Kawakami's and Maekawa's examples suggest that vowels are more likely to be deleted after fricatives, and devoiced after stops and *yō'on*, or palatalized consonants. Nihon Onsei Gakkai (1976) mentions that whether vowels are devoiced or deleted is difficult to determine when they occur word-finally.

It is not evident if vowels are devoiced in certain moras and deleted in others, even by looking at the acoustic signal. In stop consonants, the duration of aspiration period after release tends to be longer when vowels are devoiced than when they are not (see Figure 8 in section 2.8.2). However, this elongation is hard to notice in

2 This suggests the occurrence of initial geminates, which is not documented in standard Japanese. They call these [ʃʃ] and [ss] as long sounds, not as *sokuon* 'geminates' nor 'long consonants'. It is not clear if they are viewed as geminates.

fricatives, and even harder between fricatives. Vance (2008) suggests that spectral information may help the distinction, but not always.

If the vowels are deleted at the speech planning level, or phonologically, the consonants by themselves would contain no cues to the deleted vowels. However, formants peculiar to the devoiced vowel appear in the preceding consonants as mentioned in section 3.1 (Beckman and Shoji 1984; Tsuchida 1994; Faber and Vance 2001). Beckman and Shoji (1984) argue that coarticulation occurs at the speech planning level, or phonologically, before a vowel is deleted.

Durations of devoiced moras are often examined with regard to vowel devoicing. Han (1962, 1994) states that devoiced moras have similar durations to their voiced counterparts, whereas other scholars found that devoiced moras are shorter than the voiced counterparts (Beckman 1982; Hashimoto et al. 1997; Kondo 1997). If vowels are deleted in some moras and devoiced in others, the mora duration is expected to be shorter in the deleted case than in the devoiced case. However, as seen in Figure 2 (section 2.2.2), durations of devoiced moras differ considerably depending on the glottal opening pattern with which they are produced. Devoiced moras are shorter than their voiced counterparts when the glottal opening pattern for /CVC/ is reorganized into a mono-modal but it is similar in duration when the two openings are concatenated. These issues, too, require further examination.

4.5 Devoicing on accented vowels and accent shifts

Devoicing of the accented vowel diminishes the saliency of the accent which may cause listeners to fail to detect it. To avoid this inconvenience, accent tends to move to the adjacent moras when devoicing occurs on the vowel (Jinbō and Tsunemi 1932; Akinaga 1985; Sakurai 1985). This causes a systematic accent shift from HL to LH. Akinaga (1985) point out that the LH(L) pattern of the words *kisya* [kiʃa] 'steam car', *siki* [ʃiki] 'four seasons', *kikai* 'machine', and *siken* [ʃikeɴ] 'examination' (in NHK 1985) is derived from the original HL(L) pattern due to this accent shift. Moreover, the accent of the verb changed from HL to LH for *huku* 'to blow', whereas it stays HL for *kaku* 'to write' and *toru* 'to take' (Jinbō and Tsunemi 1932). In a production study, devoicing of accented vowels tended to move the accent (or pitch fall) to the following mora (Yoshida 2002). However, more recently, Akinaga (1985) observes that the pronunciation with an accent on the devoiced vowel is becoming increasingly popular among young speakers. NHK (1985) actually lists two accent patterns, LH and HL, for *kisya* [kiʃa] 'journalist' both with devoiced /i/ although it lists only one pattern, LH, for *kisya* [kiʃa] 'steam car'.

In longer words, the accent shift due to devoicing can be in the reverse direction, i.e. toward the preceding mora. The accent in the words *zinriki'sya* [ʤinrikiʃa] 'rickshaw', *gookaku'sya* [go:kakɯʃa] 'successful applicant', *koonetu'hi* [ko:netsɯçi] 'expenses for light and fuel' tends to move to the preceding mora from the original

location indicated by an apostrophe (’) if the accented vowel is devoiced. NHK (1985) lists two types of accent patterns for these words, e.g. *zinriki’sya* and *zinri’kisya*. Similarly, accent of a prefecture’s name is generally put on the mora which precedes *ken* ‘prefecture’. However, the accent of *nagasaki-ken* ‘Nagasaki Prefecture’ is either *nagasaki’-ken* or *nagasa’ki-ken*, although that of *Hirosima-ken* ‘Hiroshima Prefecture’ is uniquely *hirosima’-ken* (Bunkachō 1971). Also, the accent of city names is generally placed on the mora which precedes *si* ‘city’ as in *tatikawa’-si* ‘Tachikawa City’ and *kyooto’-si* ‘Kyoto City’, but it moves to the preceding mora when the vowel is devoiced as in *nagasa’ki-si* ‘Nagasaki City’ and *takama’tu-si* ‘Takamatsu City’ (Matsumori et al. 2012). Haraguchi (1977) analyzed the accent patterns in Japanese dialects including the accent shift due to devoicing in the frame work of autosegmental theory.

4.6 Devoicing and gemination of consonants

Consonant gemination is a phenomenon which is often discussed in connection with vowel devoicing. Gemination occurs in similar environments where devoicing occurs, as in *sentakuki* ‘washing machine’ ([sentakɯ̥ki] and [sentakki]), *gyakukooka* ‘contrary effects’ ([giakɯ̥ko:ka] and [gjakko:ka]) and *sikakukei* ‘quadrangle’ ([ʃikakɯ̥ke:] and [ʃikakke:]). Some scholars claim that gemination occurs due to vowel deletion. For example, Amanuma, Ōtsubo, and Mizutani (1993) mention that the vowel /u/ is devoiced in [sentakɯ̥ki] and deleted in [sentakki]. However, vowel deletion alone does not directly result in consonant germination. For stop and affricate gemination, deletion of oral release for the consonants is essential. That is, both oral and glottal articulation is involved in consonant gemination, whereas only glottal articulation is involved in the devoicing.

Analysis of the CSJ corpus showed that the segmental conditions are similar between consonant gemination and devoicing (Fujimoto and Kagomiya 2005). However, they differ in two crucial points. Firstly, gemination frequently occurs in a /kVk/ environment, whereas it is not the case in devoicing. The gemination of this environment includes Yamato Japanese [jokkara] from *yokokara* ‘from the side’ and [kakkoto] from *kakukoto* ‘to write’. Secondly, gemination often occurs when the preceding consonant is /r/, as in [sokkara] from *sorekara* ‘then’ and [soddewa] from *soredewa* ‘so’, which is not the general environment of devoicing (Fujimoto and Kagomiya 2005). Note that these instances as well as *yokokara* case involve the deletion of non-high vowels. These observations suggest that the mechanism that leads to consonant gemination differs from that of devoicing. Interestingly, these environments share the property with a type of verb inflection forms, or *sokuonbin*, which emerged as a historical change such as *itarite* into *itatte* ‘lead-GER’ and *torite* into *totte* ‘take-GER’ (Doi and Morita 1975).

4.7 Articulatory bases for frequent devoicing in Tokyo speakers

Frequent devoicing is a factor that characterizes Tokyo Japanese sounds as "crispy" (Sibata 1988). Traditional studies claim that frequent devoicing in this dialect stems from a general speech habit whereby strong and salient pronunciation is preferred (Sakuma 1959; Sibata 1966; Okumura 1975; Ōshima 1978; Hirayama 1985) as well as the preference of a crisply, clear speaking style (Ishiguro 1976). Consonants are pronounced more precisely and carefully than vowels in this dialect, which, in turn, makes vowels more readily devoiced (Umegaki 1968; Nihon Onsei Gakkai 1976; Mase 1977; Hirayama 1985). Moreover, consonants are claimed to be longer in the Tokyo dialect as compared to Kinki dialects (Ōno 1978; Horii 1982), which also makes vowels more readily devoiced. It is plausible that these tendencies lead to frequent devoicing in the Tokyo dialect.

Sugito (1996) reported that vowels are indeed shorter for Tokyo speakers than for Osaka speakers. Also, in /ki/ and /ke/, the consonant duration of /k/ and its VOT are relatively longer and vowels are shorter in Tokyo speakers than in Osaka speakers, whereas the closure duration of /k/ is similar across the dialects (Fujimoto et al. 2002; Fujimoto 2004a). Furthermore, glottal opening for /k/ is relatively longer and larger in the Tokyo dialect than in the Osaka dialect (Fujimoto et al. 2002). Longer aspiration (VOT) of /k/ by Tokyo speakers is manifested by earlier oral release at around the peak of glottal opening (Fujimoto et al. 2002). These results are in agreement with the above-mentioned traditional descriptions, leading to the idea that Tokyo speakers have an articulatory tendency towards devoicing.

Another finding that is characteristic of Tokyo speakers is that they may manipulate the segmental duration in a CV mora depending on vowel height. Figure 10 compares the duration of /i/ and /e/ in the /kV/ moras in non-words /kide/, /kede/ and /kete/ as a function of mora duration as produced by Tokyo and Osaka speakers (Fujimoto 2004a). Note that all are in non-general devoicing environments, and vowels are all voiced except for one /kide/ token by a Tokyo speaker, which is excluded from the analysis. In the figure, Osaka speakers are categorized into two groups: "Osaka voiced", who show no devoicing at all and "Osaka semi-devoiced", who show devoicing only in fast speech. As can be seen in the figure, vowel duration decreases proportionally as mora duration decreases in faster speech. However, the duration of /i/ by the Tokyo group is consistently short regardless of the speech rate. That is, the Tokyo speakers manipulate the mora duration by varying the consonant duration (Fujimoto 2004a). This also suggests that the Tokyo speakers control the duration of high vowels less finely in non-general devoicing environments. Vowels generally become shorter in devoiceable environments (Kondo 1997). Then, Tokyo speakers may well skip the vowels rather than producing them with even shorter voicing between two voiceless segments. Further quantitative studies are necessary to examine the relation between devoicing and the durational and articulatory characteristics in the Tokyo dialect.

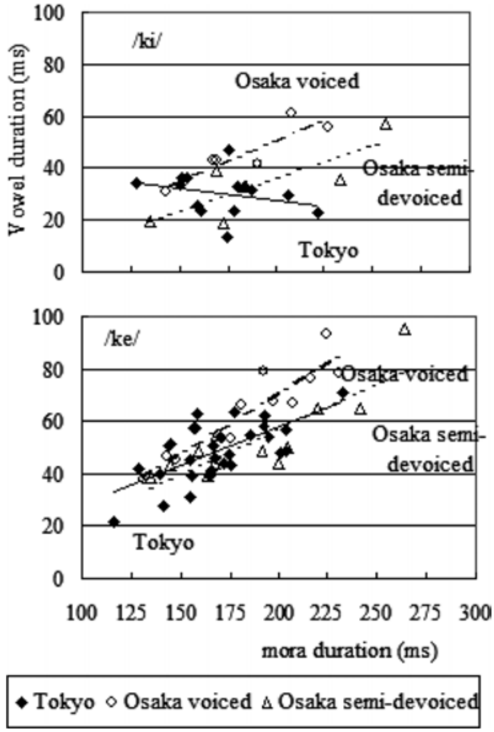


Figure 10: Vowel duration as a function of mora duration in /ki/ in /kide/ (top) and /ke/ in /kede/ and /kete/ (below), all of which are unaccented non-words. Each point represents the average of six repetitions at normal and fast speech. Regression lines for each speaker group are superimposed. (Adapted from Fujimoto 2004a)

4.8 Concatenations of glottal openings of single consonants

Acoustic studies reveal that fricatives facilitate devoicing on the following vowel more than stops do, while they tend to suppress it in post-vocalic positions (section 2.2.1). A question arises how the same consonant affects voicing and devoicing of a vowel depending on its position relative to the vowel. Also, why is devoicing more frequent in Fr-St (e.g. *sike*) than in St-Fr (e.g. *kise*)? One way to address these questions is to examine the glottal opening pattern of these consonants at different positions in the /CVC/ sequences. If the glottal openings of /s/ and /k/ merge or become mono-modal more readily in s-k than in k-s, it is likely that devoicing occurs more readily in s-k. However, such examination is difficult for Tokyo speakers, since the glottal opening pattern for Fr-St and St-Fr is both mono-modal (section 2.2.2).

Previous studies on consonant clusters in English and Swedish revealed that voiceless obstruents with aspiration or frication noise generally require a single separate glottal opening gesture on their own, while unaspirated stops tend to be

produced within the glottal opening gesture of an adjacent aspirated stop or fricative (Löfqvist and Yoshioka 1980; Yoshioka, Löfqvist, and Hirose 1981). For example, in English, the glottal opening pattern is bimodal in /s#t/ since word-initial /t/ is aspirated, whereas it is mono-modal in /#st/ since /t/ in the cluster is unaspirated (Yoshioka, Löfqvist, and Hirose 1981). The degree of glottal opening of Japanese voiceless obstruents is large in word-initial position regardless of the type of the consonant involved. In word-medial position, fricatives show large openings but stops do not (Sawashima and Niimi 1974; Fujimoto 2004b). That is, stops are more aspirated word-initially than word-medially. If we postulate that glottal openings of two neighboring consonants merge in Japanese in the same way as in English, the Fr-St sequence will more likely be produced within the same glottal opening, and St-Fr, with separate openings.

The results of an examination of an Osaka speaker support this assumption (Fujimoto 2012). Since devoicing rarely occurred for the speaker, glottal opening patterns in each consonant were observable. Figure 11 shows the glottal opening pattern of unaccented non-words /kise/ and /sike/. In the figure, the glottal opening of /kise/ shows two separate openings for /k/ and /s/ (i.e. bimodal), whereas that of /sike/ shows a salient opening for /s/ followed by a negligibly small opening for /k/ (i.e. similar to mono-modal). These patterns are alike for all 12 tokens across normal and fast speech, except for the devoiced cases.

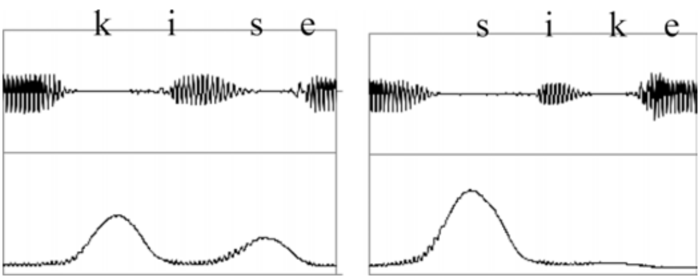


Figure 11: Speech wave (top) and glottographic signal (bottom) of voiced tokens of unaccented non-words /kise/ (left) and /sike/ (right) produced by an Osaka speaker. (Adapted from Fujimoto 2012)

Figure 12 shows the glottal opening pattern of the devoiced tokens of /kise/ and /sike/ uttered by the same speaker as in Figure 10. In /kise/, glottal openings of /k/ and /s/ concatenate into bimodal, and in /sike/ the opening is mono-modal. In another devoiced token of /kise/ by the same speaker, which is not shown here, the glottal opening has a plateau-type shape. These results suggest that glottal opening patterns become mono-modal more readily in Fr-St sequences than in St-Fr ones. This may cause the asymmetricity in devoicing rate whereby devoicing is more frequent in Fr-St than in St-Fr. Nevertheless, these data do not explain why Fr-St and Fr-St as well as St-St sequences become uniformly mono-modal in Tokyo speakers, as mentioned in section 2.2.2. Further studies are essential to answer this question.

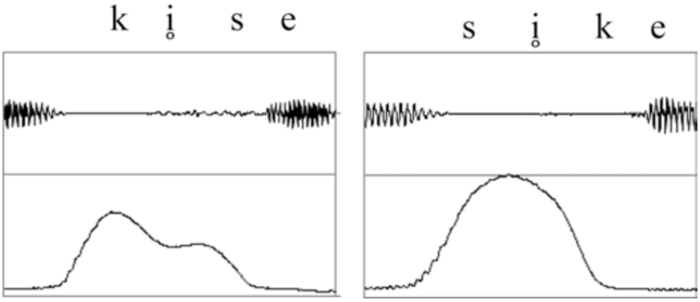


Figure 12: Speech wave (top) and glottographic signal (bottom) of devoiced tokens of unaccented non-words /kise/ (left) and /sike/ (right) produced by an Osaka speaker. (Adapted from Fujimoto 2012)

4.9 History of devoicing studies

It is not known when vowel devoicing first emerged in the history of Japanese, since the phenomenon was not described in the Japanese literature until recently (Doi and Morita 1975). Devoicing or similar phenomenon is documented in the literature of the 18th century written by foreigners. Kaempfer (1777, 1779) notes that some words lack vowels /i/ and /u/ and that many of the vowels are voiceless before consonants and in word-final position (Miyajima 1961). These conditions are identical to those of devoicing in contemporary Japanese. According to Miyajima (1961), probably the oldest description on devoicing is by Collado (1632), who pointed out that the word-final vowels are difficult to identify for beginners of Japanese. However, this condition is not quite the same as that of devoicing: "The final vowels" are not limited to high vowels. In addition, this description is similar to that of the vowel deletion phenomenon in Kagoshima in Kyushu cited by Sibata (1988), who transcribes *kaki* 'persimmon' as [kat]. Also, note that Japan closed the country from 1603 to 1867 in the Edo period, and that the habitation for foreigners was generally limited to designated areas such as Dejima Island in Nagasaki, Kyushu. Hence, as Miyajima notes, the variety of Japanese the foreigners heard and depicted is plausibly the dialects of the area they lived in.

Among Japanese scholars, Yamada (1893) referred to the effect of devoicing on accent pattern, as cited in Matsumori et al. (2012), although the phenomenon is described as "defective pronunciation" if literally translated. The term *museika* 'devoicing' is used by 1925 or earlier (Jinbō 1925, 1930; Jinbō and Tsunemi 1932). In the textbook for first-year elementary school children, Jinbō (1930) marked the moras in which the vowel would be devoiced. A detailed description on devoicing appears in Sakuma (1929, 1933, 1959). Jinbō, Sakuma, and some others may have pursued experimental phonetic studies but little is known about them. Han (1962) is the first experimental study showing spectrographic data. Sugito (1982, 1997, 1998, 2003)

made an extensive study on Japanese phonetics including devoicing. A series of physiological studies on devoicing were mostly carried out at the former Research Institute of Logopedics and Phoniatrics (RILP) at the University of Tokyo between 1965 to 1997.[3] Recently, various equipment such as EMA (electromagnetic articulography) and MRI (magnetic resonance imaging) have become available for linguistic investigations. In the physiological studies, subjects and materials are often limited due to technical, ethical, and financial reasons. However, those studies in combination will lead to a better understanding of the devoicing phenomenon.

4.10 Devoicing in other languages and L2 learners

Devoicing is reported in many other languages such as French (Schubiger 1970; Smith 2003), Montreal French (Cedergren and Simoneau 1985, cited in Beckman 1996), Korean (Jun et al. 1998a, 1998b), Spanish (Delforge 2008), Sao Miguel Portuguese (Silva 1998), Turkish (Jannedy 1994), Comanche (Armagost 1986), and Cheyenne in Algonquian, US (Milliken 1983). Schubiger (1970) lists examples in French /tant pis/ [tãpi̥] 'that's pity', /entandu/ [ãtãdy̥] 'I heared', Italian /il cane/ [il kane̥] 'the dog', and /i cani/ [i kani̥] 'the dogs', although she notes that such pronunciations are not standard. Devoicing of schwa is reported to occur in the English word *potato* (Vance 2009; Ladefoged and Johnson 2011). Ladefoged and Johnson (2011: 282) note that "it is not plausible to assume that all languages have the same set of reduction process mapping careful speech into casual speech".

Teaching devoicing to learners of Japanese has been advocated because such knowledge will help to avoid confusion in listening and to improve the naturalness of speech (Nihongo Kyōiku Gakkai 1982; Tanaka and Kubozono 1999; Imaishi 2005; Isomura 2009). Devoicing rates in L2 learners of Japanese are investigated from this pedagogical aspect. Korean learners of Japanese who live in Korea showed large inter-speaker variation and patterns in devoicing similar to those that they exhibit in their native language (Byun 2003). The devoicing rate of Taiwanese learners of Japanese who live in the Kinki area of Japan is higher in the group of advanced learners than in the group of beginners, which suggests that the devoicing skills can be acquired as their proficiency in the second language increases (Yasuda and Hayashi 2011). Also, speech rate, accent type and consonant environment seem to affect the devoicing rates of the learners who live in Kinki in much the same way as they affect the devoicing rate of the native Kinki Japanese speakers (Yasuda and Hayashi 2011).

---

3 Annual Bulletins of the Research Institute of Logopedics and Phoniatrics (RILP) are accessible on line. http://www.umin.ac.jp/memorial/rilp-tokyo/

4.11 Devoicing and its related studies

Abstract knowledge and/or articulatory gestures of devoicing seem to have a significant effect on other linguistic phenomena as well as segmental/word perception. It is well known that Japanese speakers tend to insert epenthetic vowels in consonant clusters of foreign words as in *su.to.ra.i.ku* "strike". However, if the clusters have typical consonantal conditions, speakers of the Tokyo dialect have no difficulties in producing the clusters (Tajima, Erickson, and Nagao 2000; Funatsu et al. 2008; Fujimoto and Funatsu 2008). For example, Tokyo speakers did not insert epenthetic vowels in non-words such as /apt/ and /epso/, in which the clusters have typical consonantal condition, although they did in /abt/ and /ebso/, in which the clusters have no devoicing condition (Fujimoto and Funatsu 2008). Moreover, devoiced vowels trigger more voiceless judgments than voiced vowels in the perception of the following consonants (Aoyagi and Komatsu 2003). As for the lexical representation of words, Funatsu et al. (2007) argue that words with voiced vowels are stored in the lexicon, whereas Ogasawara (2012) suggests that words with devoiced vowels may be stored. These studies are not easily comparable, since they differ in experimental settings (a brain study vs. a word shadowing study), as well as the dialectal background of the subjects (speakers from many dialects vs. Tokyo speakers only).

5 Summary

The well-known condition of vowel devoicing in Japanese is that high vowels /i/ and /u/ occur between voiceless obstruents. However, devoicing is far less frequent if both consonants are fricatives (Fr-Fr) and even less frequent if the following consonant is an /h/ or a geminate. Hence, it is reasonable to sub-categorize the general conditions into typical and atypical consonantal conditions. Typical conditions are St-St/Af, Af/Fr-St/Af, and St-Fr (non-/h/), while atypical conditions are Af/Fr-Fr, St/Af/Fr-/h/, and C̥-Vh-QC̥. In the atypical conditions, intra- and inter-speaker variations are large even among speakers of Tokyo Japanese. The distinction between typical and atypical consonantal conditions seems adequate, since the glottal opening pattern during [C̥V̥C̥] is generally mono-modal for typical conditions, and bimodal for atypical conditions.

Devoicing is also said to consistently occur between a voiceless obstruent and a pause. It is, however, generally non-systematic except for the words *desu* and *masu*, which are used very frequently in daily conversation. Devoicing occurs in non-high vowels and also in contexts where a vowel is followed by a voiced consonant. Consecutive occurrence of devoicing is documented as well. However, in these non-general environments, devoicing is usually infrequent.

Unlike schwa deletion in English, devoicing is frequent in normal speech tempo. Yet, faster speech facilitates devoicing in non-general conditions and atypical consonantal conditions. On the other hand, word accent tends to suppress devoicing with great intra- and inter-speaker variation. Dialect is a salient element which affects the devoicing rate. Kyoto-Osaka dialects are well-known for their infrequent devoicing as compared to Tokyo Japanese, but the inter-speaker variation in the former is large. Interestingly, the distinction between typical and atypical consonantal conditions holds in these dialects, too, in that devoicing is more frequent in typical ones. The effects of sociolinguistic factors such as gender and age are more subtle compared to dialectal differences. Notwithstanding the previous extensive studies, many issues have been left unsolved. This is partly because the factors contributing to devoicing are too numerous to control in building a database to analyze. Exploiting a large corpus such as CSJ will be a viable alternative.

Devoicing in Japanese shows interesting, somewhat opposing characteristics such as categorical vs. gradient nature, phonological vs. phonetic features, and devoiced vs. deleted segments. Studies that have addressed these issues are yet few. Further acoustical studies are essential to clarify the nature of devoicing in Japanese, along with further physiological studies to help elucidate the mechanism of devoicing.

References

Aoyagi, Makiko and Masahiko Komatsu. 2003. Vowel reduction and voicing judgment of following stops: Comparison between Japanese and English speakers. *Proceedings of 15th International Congress of Phonetic Sciences*, Barcelona, 1457–1460.

Akinaga, Kazue. 1985. Kyōtsūgo no akusento [Accent of standard Japanese]. In Nihon Hōsō Kyōkai (eds.), *Nihongo hatsuon akusento jiten, Kaitei shinpan* [The Japanese language pronunciation and accent dictionary, revised new edition], 128–134. Tokyo: NHK publications.

Amanuma, Yasushi, Kazuo Ōtsubo and Osamu Mizutani. 1993. *Nihongo Onseigaku* [Japanese phonetics]. Tokyo: Kurosio.

Arai, Takayuki, Natasha Warner and Steven Greenberg. 2007. Analysis of spontaneous Japanese in a multi-language telephone-speech corpus. Acoustic letter. *Acoustical Science and Technology* 28(1). 46–48.

Armagost, James. 1986. Three exceptions to vowel devoicing in Comanche. *Anthropological Linguistics* 28(3). 255–265.

Beckman, Mary E. 1982. Segment duration and the 'mora' in Japanese. *Phonetica* 39. 113–135.

Beckman, Mary E. 1996. When is a syllable not a syllable? In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 95–123. Berlin: Mouton de Gruyter.

Beckman, Mary E. and Atsuko Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41. 61–71.

Block, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Languag*e 26(1). 86–125.

Borden, Gloria J. and Katherine S. Harris. 1980. *Speech science primer: Physiology, acoustics, and perception of speech*. Baltimore/London: Lippincott Williams and Wilkins.

Browman, Catherine and Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary E. Beckman (eds.), *Papers in laboratory phonology I*, 341–376. Cambridge: Cambridge University Press.

Browman, Catherine and Louis Goldstein. 1992. Articulatory phonology. *Phonetica* 49. 155–180.

Bunkachō. 1971. *Nihongo kyōiku shidō sankōsho 1: Onsei to onsei kyōiku* [Speech and its education]. Tokyo: Ōkurashō Insatsukyoku.

Byun Hi-Gyung. 2003. Vowel devoicing in Korean and Japanese by Korean learners of Japanese. *Journal of the Phonetic Society of Japan* 7(3). 67–76.

Byun Hi-Gyung. 2007. Semaboin no museika no zenkokuteki chiikisa to sedaisa [Regional and generational differences of high vowel devoicing in Japanese]. *Nihongo no Kenkyū* 3(1). 33–48.

Byun Hi-Gyung. 2010. An age analysis of devoicing rates in five Japanese dialects. *Nihongo no Kenkyū* 6(4). 79–94.

Cedergren, Henrietta and Louise Simoneau. 1985. La chute des voyelles hautes en français de Montréal: 'As-tu entendu la belle syncope?' In Monique Lemieux and Henrietta Cedergren (eds.), *Les tendances dynamiques du français parlé à Montréal*, 57–144. Montréal Office de la Langue Française.

Collado, Diego. 1632. Ars grammaticae iaponicae lingvae. Rome: Typis & impensis Sacr. Congr. de Prop. Fide.

CSJ corpus (The Corpus of Spontaneous Japanese). 2004. National Institute for Japanese Language and Linguistics and National Institute of Information and Communications Technology. http://www.ninjal.ac.jp/english/products/csj/

Dalby, Jonathan Marler. 1986. *Phonetic structure of fast speech in American English*. Center for Speech Technology Research, University of Edinburgh. (Indiana University Linguistics Club, 1986).

Delforge, Ann Marie. 2008. Unstressed vowel reduction in Andean Spanish. In Laura Colantoni and Jeffrey Steele (eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 107–125. Somerville, MA: Cascadilla Proceedings Project.

Doi, Tadao and Takeshi Morita. 1975. *Shintei Kokugoshi Yōsetsu* [Outline of the history of Japanese, new edition]. Tokyo: Shūbunkan.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier and Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Elert, Claes-Christian. 1964. *Phonologic studies of quantity in Swedish*. Uppsala: Almqvist & Wiksell.

Faber, Alice and Timothy J. Vance. 2001. More acoustic traces of "deleted" vowels in Japanese. In Mineharu Nakayama and Charles J. Quinn, Jr. (eds.), *Japanese Korean Linguistics* 9, 100–113. Stanford: CSLI.

Fais, Laurel, Sachiko Kajikawa, Shigeaki Amano and Janet F. Werker. 2010. Now you hear it, now you don't: Vowel devoicing in Japanese infant-directed speech. *Journal of Child Language* 37(02). 319–340.

Fujimoto, Masako. 2004a. Vowel duration and vowel devoicing in Japanese: A comparison between Tokyo and Osaka dialect speakers. *Kokugogaku* 55(1). 2–15.

Fujimoto, Masako. 2004b. Effects of consonant type and syllable position within a word on vowel devoicing in Japanese. *Proceedings of the International Conference on Speech Prosody*, Nara, Japan, 625–628.

Fujimoto, Masako. 2005. Glottal opening pattern in devoiced tokens by an Osaka dialect speaker. *Journal of the Phonetic Society of Japan* 9(1). 50–59.

Fujimoto, Masako. 2006. Is vowel devoicing in Japanese phonological or phonetic?: Studies using GG, MRI, and EMG techniques. *Paper presented at Laboratory Phonology 10*, Paris, France.

Fujimoto, Masako. 2012. Effects of consonantal environment and speech rate on vowel devoicing: An analysis of glottal opening pattern. *Journal of the Phonetic Society of Japan* 16(3). 1–13.

Fujimoto, Masako, Emi Murano, Seiji Niimi and Shigeru Kiritani. 1998. Correspondence between the glottal gesture overlap pattern and vowel devoicing in Japanese. *Proceedings of 5th International Conference of Spoken Language Processing* 7, 3099–3101.

Fujimoto, Masako, Emi Murano, Seiji Niimi and Shigeru Kiritani. 2002. Difference in glottal opening pattern between Tokyo and Osaka dialect speakers: Factors contributing to vowel devoicing. *Folia Phoniatrica et Logopaedica* 54(3). 133–143.

Fujimoto, Masako and Shigeru Kiritani. 2003. Comparison of vowel devoicing for speakers of Tokyo and Kinki dialects. *Journal of the Phonetic Society of Japan* 7(1). 58–69.

Fujimoto, Masako and Takayuki Kagomiya. 2005. Gemination of consonant in spontaneous speech: An analysis of the "Corpus of Spontaneous Japanese". *IEICE Transactions on Information and Systems* E88-D(3). 562–568.

Fujimoto, Masako, Kiyoshi Honda, Niro Tayama, Ken'ichi Sakakibara and Hiroshi Imagawa. 2005. MRI and EMG studies of vowel devoicing. *Proceedings of the Autumn Meeting of Acoustical Society of Japan*, 321–322.

Fujimoto, Masako and Seiya Funatsu. 2008. Vowel epenthesis in consonant clusters by Japanese speakers. *IEICE Technical Report* SP2007-204. 105–109.

Fujimoto, Masako, Kikuo Maekawa and Seiya Funatsu. 2010. Laryngeal characteristics during the production of geminate consonants. *Proceedings of Interspeech 2010*, Makuhari, Japan, 925–928.

Fujimoto, Masako, Niro Tayama, Hiroshi Imagawa, Ken'ichi Sakakibara and Ichiro Fujimoto. 2010. Mechanism of vowel devoicing: An Osaka speaker's case. *Proceedings of the Spring Meeting of the Acoustical Society of Japan*, 359–360.

Fujimoto, Masako, Seiya Funatsu and Ichiro Fujimoto. 2012. How consonants, dialect, and speech rate affect vowel devoicing? *Proceedings of Interspeech 2012*, Portland, Oregon.

Fujisaki, Hiroya, Keikichi Hirose, Hirohumi Udagawa, Tomohiro Inoue, Tadashi Ohmori and Yasuo Satoh. 1984. Analysis of variation of variability in the acoustic-phonetic characteristics of syllables for automatic recognition of connected speech. *Transactions of the Communication on Speech Research*, The Acoustical Society of Japan S84-69. 541–548.

Funatsu, Seiya, Satoshi Imaizumi, Akira Hashizume and Kaoru Kurisu. 2007. Cortical representation of phonemic contrasts in Japanese vowel. *International Congress Series 1300*. 199–202.

Funatsu, Seiya, Satoshi Imaizumi, Masako Fujimoto, Akira Hashizume and Kaoru Kurisu. 2008. Vowel epenthesis in non-native consonant clusters. *The Bulletin of the Faculty of Human Culture and Science* 3, Prefectural University of Hiroshima. 63–71.

Funatsu, Seiya and Masako Fujimoto. 2011. Physiological realization of Japanese vowel devoicing. *Proceedings of Forum Acousticum 2011*, Aalborg, Denmark, 2709–2714.

Funatsu, Seiya, Satoshi Imaizumi, Masako Fujimoto and Ryoko Hayashi. 2011. Discrimination ability and pronunciation preference between voiced and devoiced vowels by native Japanese speakers. *Proceedings of the International Congress of Phonetic Sciences*, Hong Kong, 711–714.

Han, Mieko Shimizu. 1962. Unvoicing of vowels in Japanese. *The Study of Sounds 10* (The Phonetic Society of Japan). 81–100.

Han, Mieko Shimizu. 1994. Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96(1). 73–82.

Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.

Hashimoto, Makoto, Hiroyuki Hirai, Hidekazu Nishida and Hiroki Onishi. 1997. Analysis of relationship between devoicing vowel and mora duration. *Proceedings of the Autumn Meeting of the Acoustical Society of Japan*, 251–252.

Hasegawa, Yoko and Kazue Hata. 1992. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech* 35(1–2). 87–98.

Hattori, Shiro. 1928. Mie-ken Kameyama-chō chihō no nionsetsugo ni tsuite (1) [Two mora words in Kameyama-cho in Mie prefecture]. *Onsei Gakkai Kaihō* [The Bulletin, The Phonetic Society of Japan] 11. 11.

Hattori, Shiro. 1984. *Onseigaku* [Phonetics]. Tokyo: Iwanami.

Hibiya, Junko. 1999. Variationist sociolinguistics. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*. Massachusetts, Oxford: Blackwell. 101–120.

Hiki, Shizuo, Yoshinari Kanamori and Jiro Oizumi. 1967. On the duration of phoneme in running speech. *The Transactions of Institute of Electronics, Information, and Communication Engineers* 50(5). 69–76.

Hino, Sukezumi. 1966. Boin no museika yūseika no jittai to shojōken [Realty and conditions of vowel devoicing and voicing]. *Jinbun Ronshū* 17, Shizuoka University. 1–24.

Hirayama, Teruo. 1985. Zennippon no hatsuon to akusento [Pronunciation and accent in all Japan]. In Nihon Hōsō Kyōkai (ed.), *Nihongo hatsuon akusento jiten, Kaitei shinpan* [The Japanese language pronunciation and accent dictionary, revised new edition], 37–69. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Hirayama, Teruo. 1998. Zennippon no hatsuon to akusento [Pronunciation and accent in all Japan]. In Nihon Hōsō Kyōkai (ed.), *Nihongo hatsuon akusento jiten, Kaitei shinpan* [The Japanese language pronunciation and accent dictionary], 123–159. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Hirose, Hajime. 1971. The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica* 23. 156–170.

Hirose, Hajime. 1999. Investigating the physiology of laryngeal structures. In William Hardcastle and John Laver (eds.), *Handbook of phonetic sciences*. 116–136. Oxford: Blackwell.

Horii, Reiichi. 1982. Kinki hōgen no gaisetsu [Introduction to Kinki dialect]. In Kiichi Iitoyo, Sukezumi Hino and Ryoichi Sato (eds.), *Kōza hōgengaku 7: Kinki chihō no hōgen* [Series in dialectology 7: Dialects in Kinki region], 1–25. Tokyo: Kokusho Kankōkai.

Imai, Terumi. 2004. *Vowel devoicing in Japanese*. Michigan: Michigan State University dissertation.

Imai, Terumi. 2010. An emerging gender difference in Japanese vowel devoicing. In Dennis R. Preston and Nancy Niedzielski (eds.), *A reader in sociophonetics*, 177–187. New York: De Gruyter Mouton.

Imaishi, Motohisa. 2005. *Onsei kenkyū nyūmon* [Introduction to phonetic studies]. Osaka: Izumi Shoin.

Imaizumi, Satoshi, Akiko Hayashi and Toshisada Deguchi. 1995. Listener adaptive characteristics of vowel devoicing in Japanese dialogue. *Journal of the Acoustical Society of America* 98(2). 768–778.

Imaizumi, Satoshi, Kiyoko Fuwa and Hiroshi Hosoi. 1999. Development of adaptive phonetic gestures in children: Evidence from vowel devoicing in two different dialects of Japanese. *Journal of the Acoustical Society of America* 106(2). 1033–1044.

Inoue, Fumio. 1968. *Tōhoku hōgen no shiin taikei* [Consonant system in Tohoku dialects]. *Gengo Kenkyu* 52. 80–98.

Ishiguro, Rohei. 1976. *Tōkyō hōgen on'inkō* [Phonology of the Tokyo dialect]. In Hidekatsu Saitoh (ed.), Tōkyō hōgenshū, 343–389. Tokyo: Kokusho Kankōkai.

Isomura, Kazuhiro. 2009. *Kokusai kōryū kikin nihongo kyōjuhō shirīzu 2: Onsei o oshieru* [Series in Japanese language pedagogy 2: Teaching speech sounds, The Japan Foundation]. Tokyo: Hituzi Syobo.

Jannedy, Stefanie. 1994. Prosodic and segmental influences on high vowel devoicing in Turkish. *Proceedings of the Fifth Australian International Conference on Speech Science and Technology*, 674–679.

Jinbō, Kaku. 1925. *Kokugo Onseigaku* [Japanese phonetics]. Tokyo: Meiji Shoin.

Jinbō, Kaku. 1930. *Kokugo dokuhon no hatsuon to akusento: Jinjō ichi gakunen* [Pronunciation and accent of the Japanese reader: First grade elementary school]. Tokyo: Kōseikaku.

Jinbō, Kaku and Senri Tsunemi. 1932. *Kokugo hatsuon akusento jiten: Kaisetsu* [Japanese pronunciation and accent dictionary: Commentary]. Tokyo: Kōseikaku.

Jun, Sun-Ah and Mary E. Beckman. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the Annual Meeting of the Linguistic Society of America, Los Angeles, 7–10 January.

Jun, Sun-Ah, Mary E. Beckman, Seiji Niimi and Mark Tiede. 1998a. Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *UCLA Working Papers in Phonetics* 96. 1–42.

Jun, Sun-Ah, Mary E. Beckman and Hyuck-Joon Lee. 1998b. Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. *UCLA Working Papers in Phonetics* 96. 43–68.

Japanese Language Council. 1954. *Hyōjungo no tameni* [Towards standard Japanese].

Kaempfer, Engelbert. 1777–1779. *Geschichte und Beschreibung von Japan*. Lemgo, im Verlage der Meyerschen Buchhandlung.

Kagaya, Ryuhei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics* 2. 161–180.

Kawai, Hisashi, Norio Higuchi, Tohru Shimizu and Sei'ichi Yamamoto. 1995. Devoicing rules for text-to-speech synthesis of Japanese. *Journal of the Acoustical Society of Japan* 51(9). 698–705.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawakami, Shin. 1969. Musei haku no tsuyosa to akusento kaku [Intensity of devoiced mora and accent nucleus]. *Kokugo Kenkyū* 27. Reprinted in Shin Kawakami 1995. *Nihongo akusento ronshū* [Papers on Japanese accent], 316–331. Tokyo: Kyūko Shoin.

Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to the sounds of Japanese]. Tokyo: Ōfūsha.

Kawatsu Hiromi and Kikuo Maekawa. 2009. Influence of the manner of articulation on vowel devoicing rate: Analysis of the Corpus of Spontaneous Japanese. *Proceedings of the Spring Meeting of the Acoustical Society of Japan*, 443–444.

Kimura, Osamu, Nobuyoshi Kaiki and Atsunori Kito. 1998. Analysis of vowel devoicing rules for synthesis-by rule. *Proceedings of the Spring Meeting of the Acoustical Society of Japan*, 137–138.

Kindaichi, Haruhiko. 1954. On'in [Phonology]. In Misao Tōjō (ed.), *Nihon hōgengaku* [Japanese dialectology], 87–176. Tokyo: Yoshikawa Kōbunkan.

Kindaichi, Haruhiko and Kazue Akinaga. 2001. Kaisetsu [Commentary]. In *Shinmeikai nihongo akusento jiten* [Shinmeikai Japanese accent dictionary]. Tokyo: Sanseido.

Kitahara, Mafuyu and Shigeaki Amano. 2001. Perception of pitch accent categories in Tokyo Japanese. *Gengo Kenkyu* 120. 1–34.

Kiritani, Shigeru, Hiroshi Imagawa and Hajime Hirose. 1996. Vocal cord vibration in the production of consonants: Observation by means of high-speed digital imaging using a fiberscope. *Journal of the Acoustical Society of Japan (E)17*. 1–8.

Komatsu, Masahiro and Makiko Aoyagi. 2005. Vowel devoicing vs. mora-timed rhythm in spontaneous Japanese: Inspection of phonetic labels of OGI_TS. *Proceedings of Interspeech 2005*, Lisbon, Portugal, 2461–2464.

Kondo, Mariko. 1997. Mechanisms of vowel devoicing in Japanese. Edinburgh: University of Edinburgh dissertation.

Kondo, Mariko. 2001. Vowel devoicing and syllable structure in Japanese. In Mineharu Nakayama and Charles J. Quinn, Jr. (eds.), *Japanese/Korean linguistics 9*, 125–138. Stanford: CSLI.

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing environments. In Jeroen Maarten van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 229–245. Berlin and New York: Mouton de Gruyter.

Kuriyagawa, Fukuko and Masayuki Sawashima. 1986. Vowel duration in Japanese /t̪su̥ku/ and /t̪suku̥/. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 20. 119–130.

Ladefoged, Peter and Keith Johnson. 2011. *A course in phonetics*, 6th edition. Boston: Wadsworth.

Lehiste, Ilse. 1977. *Suprasegmentals*. Cambridge: MIT Press.

Lisker, Leigh, Arthur S. Abramson, Franklin S. Cooper and M. H. Schvey. 1969. Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45. 1544–1546.

Löfqvist, Anders and Hirohide Yoshioka. 1980. Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 68(3). 792–801.

Löfqvist, Anders and Hirohide Yoshioka. 1981. Interarticulator programing in obstruent production. *Phonetica* 38. 21–34.

Maekawa, Kikuo. 1983. Kyōtsūgo ni okeru boin no museika no kakuritsu ni tsuite [Probability of devoicing in standard Japanese] in *Gengo no Sekai* 1–2 [World of language], 69–81. Reprinted in *Kokugogaku Ronsetsu Shiryō* 20(5). 71–77.

Maekawa, Kikuo. 1989. Boin no museika [Devoicing of Vowels]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 178–205. Tokyo: Meiji Shoin.

Maekawa, Kikuo. 1990. Hatsuwa sokudo ni yoru yūsei kukan no hendō [Effects of speaking rate on the voicing variation in Japanese]. *IEICE Technical Report* SP89-148. 47–53.

Maekawa, Kikuo. 2011. Kōpasu o riyō shita jihatsu onsei no kenkyū [A phonetic study on spontaneous Japanese using corpus]. Tokyo: Tokyo Institute of Technology dissertation.

Maekawa, Kikuo. this volume. Chapter 16: Corpus-based phonetics. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Maekawa, Kikuo and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In Jeroen Maarten van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 205–228. Berlin and New York: Mouton de Gruyter.

Mase, Yoshio. 1977. Tōzai hōgen no tairitsu [The contrast between eastern and western dialects]. In Susumu Ōno and Takesi Sibata (eds.), *Iwanami kōza nihongo 11: Hōgen* [Iwanami Japanese series 11: Dialects], 235–289. Tokyo: Iwanami.

Matsumori, Akiko, Tetsuo Nitta, Yoko Kibe and Yukihiko Nakai. 2012. *Nihongo akusento nyūmon* [Introduction to Japanese accent]. Tokyo: Sanseido.

McCawley, James. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Milliken, Stuart. 1983. Vowel devoicing and tone recoverability in Cheyenne. *Working papers of the Cornell phonetics laboratory* 1. 43–76.

Miyajima, Tatsuo. 1961. Boin no museika wa itsu kara atta ka [When did vowel devoicing start?]. *Kokugogaku* 45. 38–48.

Morris, Midori Yonezawa. 2010. Perception of devoicing variation and the judgment of speakers' region in Japanese. In Dennis R. Preston and Nancy Niedzielski (eds.), *A reader in sociophonetics*, 192–202. New York: De Gruyter Mouton.

Munhall, Kevin and Anders Löfqvist. 1992. Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20. 111–126.

Nagano-Madsen, Yasuko. 1994a. Vowel devoicing rates in Japanese from a sentence corpus. *Lund Working Papers in Linguistics* 42. 117–127.

Nagano-Madsen, Yasuko. 1994b. Influence of accent and tone on the realization of vowel devoicing in Japanese: Analysis of a sentence database. *Lund Working Papers in Linguistics* 43. 104–107.

Nagano-Madsen, Yasuko. 1995. Effect of accent and segmental contexts on the realization of vowel devoicing in Japanese. *Proceedings of 13th International Conference of Spoken Language Processing 95*(3), 564–567.

Nakao, Toshio, Junko Hibiya and Noriko Hattori. 1997. *Shakai gengogaku gairon: Nihongo to eigo no rei de manabu shakai gengogaku* [Introduction to sociolinguistics: Sociolinguistics learning with Japanese and English examples]. Tokyo: Kurosio.

NHK (Nihon Hōsō Kyōkai). 1985. *NHK nihongo akusento jiten, kaiteiban* [Dictionary of pronunciation and accent in Japanese: Revised edition]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

NHK (Nihon Hōsō Kyōkai). 2005. *NHK anaunsu jissen toreiningu* [Practical training of NHK broadcast announcers]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Nihon Onsei Gakkai. (ed.). 1976. *Onseigaku daijiten* [Dictionary of phonetics]. Tokyo: Sanshūsha.

Nihongo Kyōiku Gakkai (ed.). 1982. *Nihongo kyōiku jiten* [Dictionary of teaching Japanese]. Tokyo: Taishukan.

Ogasawara, Naomi (2012). Lexical representation of Japanese vowel devoicing. *Language and Speech* 56(1). 5–22.

Okumura, Mitsuo. 1975. Kinki no hōgen [Dialects in Kinki]. In Hatsutaro Oishi and Yukio Uemura (eds.), *Hōgen to hyōjungo: Nihongo hōgengaku gaisetsu* [Dialects and standard language: Introduction to Japanese dialectology], 264–294. Tokyo: Chikuma Shobō.

Ōno, Susumu. 1978. Nihon no tōbu to seibu [Eastern and western Japan]. In Takesi Sibata, Masanobu Kato and Munemasa Tokugawa (eds.), *Nihon no gengogaku 6: Hōgen* [Japanese linguistics 6: Dialectology], 100–119. Tokyo: Taishukan.

Ōshima, Masatake. 1978. Chiiki hatsuon no henka oyobi sono haifu [Changes and the propagation of the pronunciation of rural areas]. In Takesi Sibata, Masanobu Kato and Munemasa Tokugawa (eds.), *Nihon no gengogaku 6: Hōgen* [Japanese linguistics 6: Dialectology], 75–82. Tokyo: Taishukan.

Ohso, Mieko. 1973. A phonological study of some English loan words in Japanese. *Ohio State University Working Papers in Linguistics* 14. 1–26.

Sagisaka, Yoshinori and Yoh'ichi Tohkura. 1984. Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyo [Phoneme duration control for speech synthesis by rule]. *Denshi Tsūshin Gakkai Ronbunshi* [The Transactions of the Institute of Electronics, Information and Communication Engineers A] 67(7). 629–636.

Sakuma, Kanae. 1929. *Nihon onseigaku* [Japanese phonetics]. Tokyo: Kyōbunsha. Reprinted in 1963. Tokyo: Kazama Shobō.

Sakuma, Kanae. 1933. *Kokugo onseigaku gaisetsu* [Introduction to Japanese phonetics]. Tokyo: Dobunkan.

Sakuma, Kanae. 1959. *Hyōjun nihongo no hatsuon akusento* [Pronunciation and accent in standard Japanese]. Tokyo: Kōseikaku.

Sakurai, Shigeharu. 1966. Kyōtsūgo no hatsuon de chūi subeki kotogara [Matters requiring attention in the pronunciation of standard Japanese]. In NHK (ed.), *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary], 31–43. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Sakurai, Shigeharu. 1985. Kyōtsūgo no hatsuon de chūi subeki kotogara [Matters requiring attention in the pronunciation of standard Japanese]. In NHK (ed.), *Nihongo hatsuon akusento jiten, Kaitei shinpan* [The Japanese language pronunciation and accent dictionary, revised new edition], 128–134. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Sawashima, Masayuki. 1969. Vowel devoicing in Japanese: A preliminary study by photoelectric glottography. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 3. 35–41.

Sawashima, Masayuki. 1971a. Use of fiberscope and observing articulatory movements. *Journal of the Acoustical Society of Japan* 27(9). 425–434.

Sawashima, Masayuki. 1971b. Devoicing of vowels. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 5. 7–13.

Sawashima, Masayuki and Sachio Miyazaki. 1973. Glottal opening for Japanese voiceless consonants. *Annual Bulletin of Research Institute Logopedics and Phoniatrics* 7. 1–9.

Sawashima, Masayuki and Seiji Niimi. 1974. Laryngeal conditions in articulations of Japanese voiceless consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 8. 13–17.

Sawashima, Masayuki, Hajime Hirose and Hirohide Yoshioka. 1978. Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 12. 53–60.

Schubiger, Maria. 1970. *Einfürung in die Phonetic* [Introduction to phonetics]. Walter de Gruiter. (Japanese translation by Tamotsu Koizumi. 1973. *Onseigaku nyūmon* [Introduction to phonetics]. Tokyo: Taishukan.)

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shrosbree, Miki. 2013. A study of a blocking factor of Japanese devoicing: Comparison of pre-geminate vs. pre-singleton consonant environments. *Phonological Studies* 16. 61–68.

Sibata, Takesi. 1966. Museika [Devoicing]. In The Society for Japanese Linguistics (ed.), *Kokugogaku jiten* [Glossary of Japanese linguistics], 899. Tokyo: Tōkyōdō Shuppan.

Sibata, Takesi. 1988. *Hōgenron* [Dialectology]. Tokyo: Heibonsha.

Silva, David James. 1998. Vowel lenition in São Miguel Portuguese. *Hispania* 81. 166–178.

Smith, Caroline. 2003. Vowel devoicing in contemporary French. *Journal of French Language Studies* 13(2). 177–194.

Sugito, Miyoko. 1982. *Nihongo akusento no kenkyū* [Study on Japanese accent]. Tokyo: Sanseido.

Sugito, Miyoko. 1996. *Nihongo onsei no kenkyū 3: Nihongo no oto* [Study of Japanese prosody and speech sounds 3: Japanese sounds]. Osaka: Izumi Shoin.

Sugito, Miyoko. 1997. *Nihongo onsei no kenkyū 4: Onsei hakei wa kataru* [Study of Japanese prosody and speech sounds 4: What speech waveforms tell us]. Osaka: Izumi Shoin.

Sugito, Miyoko. 1998. *Nihongo onsei no kenkyū 6: Shibata-san to Imada-san* [Study of Japanese prosody and speech sounds 6: Mr. Shibata and Mr. Imada]. Osaka: Izumi Shoin.

Sugito, Miyoko. 2003. Timing relationships between prosodic and segmental control in Osaka Japanese word accent. *Phonetica* 60. 1–16.

Sugito, Miyoko and Hajime Hirose. 1988. Production and perception of accented devoiced vowels in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 22. 21–39.

Tajima, Keiichi, Donna Erickson and Kyoko Nagao. 2000. Factors affecting native Japanese speakers' production of intrusive (epenthetic) vowels in English words. *Proceedings of the 2000 International Conference on Spoken Language Processing*, Beijing, China, 585–561.

Takeda, Kazuya and Hisao Kuwabara. 1987. Boin museika no yōin bunseki to yosoku shuhō no kentō [Analysis and perception of devocalizing phenomena]. *Proceedings of the Autumn Meeting of the Acoustical Society of Japan*, 105–106.

Tanaka, Shin'ichi and Haruo Kubozono. 1999. *Nihongo no hatsuon kyōshitsu: Riron to renshū* [Introduction to Japanese pronunciation: Theory and practice]. Tokyo: Kurosio.

Tsuchida, Ayako. 1994. Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working papers of the Cornell Phonetic Laboratory* 9. 183–222.

Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. Ithaca: Cornell University dissertation.

Tsuchida, Ayako. 2001. Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10(3). 225–245.

Umegaki, Minoru. 1968. Kinki hōgen [Dialects in Kinki]. In Japanese Linguistic Society (ed.), *Hōgengaku gaisetsu* [Introduction to dialectology], 82–86. Tokyo: Musashino Shoin.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.

Varden, John Kevin. 2010. On vowel devoicing in Japanese. *Karuchuru* (Journal of Liberal Arts Studies) 4(1). 223–235. Meiji Gakuin University.

Yamada, Bimyo. 1893. Nihon onchōron [Japanese accentology]. Supplement of *Nihon Daijisho* [Dictionary of Japanese], 43–57. Tokyo: Nihon Daijisho Hakkōsho.

Yasuda, Rei, and Ryoko Hayashi. 2011. Acquisition of vowel devoicing in Japanese as a second language by Taiwanese learners of Japanese. *Journal of the Phonetic Society of Japan* 15(2). 1–10.

Yoshida, Natsuya. 2002. Onsei kankyō ga boin no museika ni ataeru eikyō ni tsuite [The effects of phonetic environment on vowel devoicing in Japanese]. *Kokugogaku* 53(3). 34–47.

Yoshida, Natsuya. 2008. /p/ ni kōzoku suru museika boin o chōkaku suru onkyōjō no tegakari [The acoustic cue of the perception of devoiced vowels after /p/]. *Journal of the Phonetic Society of Japan* 12(3). 52–58.

Yoshida, Natsuya and Yoshinori Sagisaka. 1990. Factor analysis of vowel devoicing in Japanese. *ATR Technical Report* TRI-0159. 1–9.

Yoshioka, Hirohide. 1981. Laryngeal adjustment in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38. 236–351.

Yoshioka, Hirohide, Anders Löfqvist and Hajime Hirose. 1981. Laryngeal adjustment in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America* 70(6). 1615–1623.

Yoshioka, Hirohide, Anders Löfqvist and Hajime Hirose. 1982. Laryngeal adjustment in Japanese voiceless sound production. *Journal of Phonetics* 10. 1–10.

Yamamoto, Toshiharu. 1982. Osaka-fu no hōgen [Dialects in Osaka Prefecture]. In Sukezumi Hino, Kiichi Iitoyo and Ryoichi Sato (eds.), *Kōza hōgengaku 7: Kinki chihō no hōgen* [Series in dialectology 7: Dialects in Kinki region], 195–200. Tokyo: Kokusho Kankōkai.

Haruo Kubozono

5 Diphthongs and vowel coalescence

1 Introduction

In this chapter we will tackle the following three questions concerning diphthongs in modern Tokyo Japanese: (i) what constitutes diphthongs as against heterosyllabic vowel sequences (vowel hiatus), (ii) why and how /ai/ constitutes a stable diphthong, whereas /au/ fails to do so, and (iii) what is the rule that provides a principled account of the various patterns of vowel coalescence whereby vowel sequences of different qualities turn into monophthongs, e.g., /ai/ → /e/, /eu/ → /o/. We will tackle the first question by examining the behavior of various vowel sequences in accentual and other phenomena to show that /ai/, /oi/ and /ui/ are real diphthongs in modern Japanese. As for the second question, we will account for various facts pertaining to diphthongs from synchronic, diachronic, cross-linguistic and phonetic perspectives to substantiate the observation that /ai/, /oi/ and /ui/ are stable diphthongs as against /au/, /eu/, /iu/ and other vowel sequences. The third question will be tackled by examining the various seemingly complex patterns of vowel coalescence in terms of phonetic features.

This chapter is organized as follows. In the next section, we will first define the term "diphthongs" and then look at some accentual and other phenomena to determine which vowel sequences constitute a diphthong in modern Japanese. Section 3 develops this argument to demonstrate that /ai/ and /au/ exhibit asymmetries in the language. We will examine several independent phenomena to understand that /au/ resists being fused into a diphthong. Section 4 widens our scope beyond Japanese to understand that /ai/-/au/ asymmetries are also observed in English and several other languages. This will be followed by a discussion in Section 5 where phonetic reasons for the asymmetries are considered. Section 6 constitutes an independent section where different patterns of vowel coalescence are analyzed. It is in this chapter that we consider a rule that can generalize seemingly complex patterns of vowel coalescence. The final section (section 7) gives a summary of the chapter and the major issues that are to be solved in the future.

2 Diphthong

2.1 Definition

Diphthongs refer to a tautosyllabic sequence of two vowels of different qualities. One question that always arises when we discuss this type of vowel is where we can draw a line between "diphthongs" as defined in this way and heterosyllabic vowel

sequences known as "hiatus", or sequences of adjacent vowels that occur across a syllable boundary (Casali 2011).

At least three criteria are used for the definition of diphthongs in general, one being morphological and the other two phonological. The morphological criterion is that two vocalic elements must be within a morpheme rather than across two morphemes in order to form a diphthong. Thus /ai/ in the word /ai/ 'love' is entitled to form a diphthong, whereas /ai/ in the compound noun /ha+isya/ 'tooth, doctor; dentist' is not. Of the two phonological criteria, one concerns the sonority of the two vowels in question. Given a sequence of two vowels, V_1 and V_2, V_1 must be at least as sonorous as V_2 to form a diphthong. Stated conversely, V_1 and V_2 belong to different syllables and do not form a diphthong if V_2 is more sonorous than V_1, e.g., /ia/, /oa/. Potential exceptions to this are cases where the first vowel becomes a glide, e.g., /ia/ → /ja/, as well as cases where the second vowel becomes a schwa, e.g., /ia/ → /iə/.[1]

The other phonological criterion is related to word accent and specifically applies to Japanese. The accent assigned to V_2 by any quantity-sensitive accent rule usually shifts to V_1 if the two vowels are within a single syllable, i.e., if they form a diphthong. In contrast, the accent assigned to V_2 remains intact if the two vowels are across a syllable boundary. This interpretation is based on the general observation that accent falls on the nuclear vowel of the syllable, rather than on the mora originally designated as the accent locus by mora-counting accent rules (McCawley 1978; Kubozono 1999a).

Although matters may be more complicated in some cases, the three criteria stated above seem sufficient when we discuss vowel sequences in Japanese. Generally, both /ai/ and /au/ satisfy the two phonological requirements as long as the two vowels are within a single morpheme. Other vowel sequences such as /oi/, /ei/, /eu/ and /ou/ can also be interpreted as constituting a diphthong as long as they are tautomorphemic. /iu/ and /ui/ may be somewhat more ambiguous because their components, /i/ and /u/, are just as sonorous as each other (Selkirk 1984). These vowel sequences must be tested by accent rules with respect to their syllabic status.

2.2 Diphthongs in Kagoshima Japanese

As mentioned above, there are two types of vowel sequences in language in general, those that constitute a diphthong (or a long vowel) and those that do not. By definition, the first type forms a single syllable, whereas the second type forms two

1 Modern Japanese permits sequences like /ja/, /ju/, /jo/ and /wa/: e.g., /ja.ku/ 'role', /kjaku/ 'guest', /kjuu/ 'nine', /mjoo/ 'strange', /wa/ 'peace'. In this chapter, these are regarded as onset-nucleus sequences rather than diphthongs, or vowel-vowel sequences in the nucleus, because they behave quite differently from /ai/ and other real diphthongs in Japanese phonology: e.g., /ja/ counts as one mora, with /j/ not contributing to syllable weight, whereas /ai/ counts as two moras.

syllables. In a mora-based dialect like Tokyo Japanese and Kyoto/Osaka Japanese, it is quite difficult to tell whether a certain vowel sequence forms one unified syllable or two separate syllables (see Kubozono 1999a, 2001 for details). Not surprisingly, different scholars have posited different vowel sequences as diphthongs in the literature. For example, Kawakami (1977) suggested that /ae/, /ao/ and /oe/ as well as /ai/, /oi/ and /ui/ may form a diphthong in Japanese, while Saito (1997) assumed that /ae/ and /au/ constitute diphthongs as do /ai/, /oi/ and /ui/. Kibe (2000) mentions specifically for Kagoshima Japanese that /ai/, /oi/ and /ui/ are diphthongs in this dialect although she does not provide negative evidence for other vowel sequences.

While the syllable structure of a certain vowel sequence cannot be easily determined in Tokyo and other mora-counting dialects of Japanese, this can be done quite readily in Kagoshima Japanese, where accent (tone) assignment rules directly refer to syllable structure. This dialect has two accent patterns, often called Type A and Type B (Hirayama 1951), both of which have one and only one high-toned *syllable* per word. Specifically, Type A has a high tone on the penultimate syllable, whereas Type B bears a high tone on the final syllable. This system thus enables us to judge the syllable structure of words, especially those that contain a vowel sequence (Kubozono 2004).

Let us consider the case of /oe/, first of all. An examination of its accentual behavior in Kagoshima Japanese shows that /o/ and /e/ are assigned different tones in relevant positions, which, in turn, means that /oe/ behaves as a heterosyllabic vowel sequence. The bimoraic morpheme /koe/ 'voice', for example, bears a high tone on /e/ when pronounced in isolation: i.e., /koE/ (capital letters denote a high-toned portion). This itself suggests that /e/ forms an independent syllable on its own, i.e., /ko.E/, just as the bimoraic word /ko.ME/ 'rice' is high toned on the final syllable.

This analysis can be further confirmed by the tonal pattern of compound nouns of which /koe/ forms a second member. In Kagoshima Japanese, compound words and phrases inherit the accent pattern of their initial member, so that this tonal pattern spreads over the domain of the entire compound. Thus, compound nouns with an A-type initial member have a high tone on their penultimate syllable, while those with a B-type initial member are high-toned on the final syllable. When /koe/ 'voice' is combined with an A-type morpheme like /uta/ 'song' to form the compound noun /uta-goe/[2] 'singing voice', /go/ is pronounced with a high tone, i.e., /u.ta-GO.e/. This indicates that /go/ alone constitutes the penultimate syllable of the word. Furthermore, when combined with a B-type morpheme like /oo/ 'big', /e/ is high-toned, indicating that this is the final syllable of the word. This whole analysis is summarized in (1), where /-/ denotes a morpheme boundary.

2 /koe/ undergoes *rendaku*, or sequential voicing, by which the initial consonant of the second member of the compound is voiced: Vance (this volume) for more details about this morphophonological process.

(1) a. ko.E ‘voice, sound’
b. u.ta-GO.e ‘singing voice’ (/u.ta/ being an A-type morpheme)
c. oo-go.E ‘big voice’ (/oo/ being a B-type morpheme)

On the other hand, /oi/ exhibits a diphthong-like behavior in the same dialect. For example, the word /koi/ ‘carp’ is pronounced with a high tone entirely, i.e., /KOI/, suggesting that it is a monosyllabic B-type morpheme just like /TOO/ ‘tower’ and /MON/ ‘gate’. This analysis has been confirmed by the fact that the compound noun /ma-goi/ ‘black carp’ bears a high tone on /goi/, i.e., /ma-GOI/.

Kubozono (2004) analyzed the syllable structure of every possible vowel sequence by combining the five vowels of Japanese. Sequences of an identical vowel were excluded since they form a long vowel and, hence, a single syllable, as long as they occur within a single morpheme. In addition, /ei/ and /ou/ were also excluded from analysis since they tend to be pronounced /ee/ and /oo/, respectively, in normal speech. Moreover, /iu/ and /eu/ do not seem to exist in the morphemes of modern Japanese probably because they turned into /juu/ and /joo/ in the history of the language (section 3.2). This leaves sixteen combinations of vowels. The actual words that were tested fall into three kinds according to their origin: (a) native Japanese words, (b) Sino-Japanese (SJ) words, and (c) loanwords other than those from Chinese.

Analysis of these words with respect to their accentual behavior has confirmed Kibe’s (2000) idea. It revealed that only three vowel sequences out of sixteen form a diphthong in Kagoshima Japanese, namely, /ai/, /oi/ and /ui/, whose accentual behavior is shown in (2)–(4). (a)–(c) correspond to the three kinds of words mentioned above. Note that most loanwords belong to the A-type in this dialect, i.e., bear a high tone on the penultimate syllable. /oi/ is systematically absent in SJ morphemes of modern Japanese.

(2) /ai/
a. native word
KAI ‘shellfish’, a.KA-gai ‘arch shell’, ni.mai-GAI ‘bivalve, lamellibranch’; TAI ‘sea bream’, a.KA-dai ‘red sea bream’, ku.ro-DAI ‘black sea bream’
b. SJ word
han.TAI ‘opposition, objection’, TAI.sen ‘match-up’, sa.WA-kai ‘tea party’
c. Loanword
ne.ku.TAI ‘necktie’, MA.sai ‘The Masais (tribe)’, ee.AI ‘AI, artificial intelligence’

(3) /oi/
a. ni.OI ‘smell’, KOI ‘carp’, ma-GOI ‘black carp’, ni.wa.TOI ‘chicken’
b. N.A.
c. MOI.ra ‘Moira (girl’s name)’, ROI.do ‘Lloyd (family name)’

(4) /ui/

a. ku.SUI 'medicine', kaze-GU.sui 'medicine for a cold', KE.mui 'smoke'
sui.KA 'water melon', ki.i.ro-ZUI.ka 'yellow water melon'
oo-GUI 'big eater', TUI.ni 'finally'

b. IT-tui 'one pair', SAN-tui 'three pairs', ge.KI-tui 'shoot down', SUI-yoo 'Wednesday'

c. KUI.zu 'quiz', SUI.su 'Swiss', ee.BUI 'AV, audio-visual'

While the three vowel sequences given in (2)–(4) function as a diphthong, the other thirteen vowel sequences behave like a heterosyllabic vowel sequence rather than a diphthong. These include sequences involving a falling sonority as in (5), those involving a rising sonority as in (6), and those that involve neither a rise nor a fall in sonority as in (7). Vowel sequences involving a sonority rise are predominantly found in loanwords.

(5) /au/ do.NA.u 'Donau, Danube', do.na.U-ga 'Donau-NOM', san.pa.U.ro 'Sao Paulo'

/ao/ a.O 'blue', mas-SA.o 'pale face', KA.o 'face', ma-ga.O 'sober face', a.sa-ga.O 'morning glory (flower)', ta.O.ru 'towel'

/ae/ a.ka-ga.E.ru 'red frog', ki-GA.e 'change of clothes', ka.E.de 'maple'

(6) /ea/ wan-PE.a 'one pair', tuu-pe.A 'two pairs', a.hu.taa-KE.a 'after(-sales) care'

/oa/ ko.A.ra 'koala bear', si.ro-ko.a.RA 'white koala bear', DO.a 'door'

/ia/ RI.a 'Lear', ri.A-oo 'King Lear', pi.A.no 'piano', in.do.ne.SI.a 'Indonesia'

/io/ si.O 'salt', o-si.O 'salt (polite form)', pi.no.KI.o 'Pinocchio'

/ie/ i.E 'house', i.e-i.E 'a block of houses', oo-mi.E 'chest thumping'

/ua/ a.KU.a 'aqua', RU.aa 'lure', TU.aa 'tour'

/uo/ U.o 'fish', u.O-tui 'fishing', a.ka-U.o 'red fish'

/ue/ U.e 'upper (side)', u.E-no 'Ueno (family name)', hu.E 'flute', ta-u.E 'rice planting'

(7) /eo/ bi.DE.o 'video', bi.de.O-ten 'video shop', NE.on 'neon', ne.ON-gai 'neon-lit street'

/oe/ a.RO.e 'aloe', a.ro.E-syoo 'aloe dealer', oo-go.E 'big voice', ya.ma-zo.E 'Yamazoe (personal name)'

The data in (2)–(7) reveal that Kagoshima Japanese has only three diphthongs: /ai/, /oi/ and /ui/. Interestingly, these vowel sequences all end in /i/. Since /ei/

alternates with /ee/ and functions as one syllable in Kagoshima Japanese just as it does in other dialects, it follows that all vowel sequences that end in /i/ are qualified as a diphthong. All other types of vowel sequence split into two syllables in this dialect.

This result itself may not be very surprising since /i/ is a high vowel and vowel sequences ending in this vowel involve a high degree of sonority decline. It is well known in general phonology that a consonant sequence forms a more stable and well-formed consonant cluster in both onset and coda positions as it involves a greater sonority difference (Selkirk 1984): For example, /pl/ forms a better onset than /ml/ or /pm/ cross-linguistically. If the same principle applies to complex nuclei, i.e., vocalic clusters within a single syllable, it can be supposed that a vowel sequence forms a better and more stable diphthong as it involves a higher degree of sonority decline. Given this idea, /ai/ should be one of the most stable diphthongs cross-linguistically, followed possibly by /oi/. On the other hand, it is quite strange to find that /ui/ forms a diphthong in Kagoshima Japanese since it consists of two vowels of a roughly equal sonority rank.

More peculiar is the fact that vowel sequences ending in /u/ do not constitute a diphthong at all. /eu/ and /iu/ turned into /joo/ and /juu/ in the history of the language, as we will see shortly below, and do not seem to exist in any Japanese morpheme (except possibly in certain new loanwords). However, it remains unclear why /au/ forms two separate syllables rather than one. That /ai/ but not /au/ functions as a diphthong in the phonological system of Kagoshima Japanese can be fully substantiated by the data in (8).[3]

(8) a. /ai/
FAI.ru 'file', a.ri.BAI 'alibi', MA.sai 'The Masais', a.ru.BAI.to 'Arbeit, part-time work', PAI.ron 'Pairon (name of a medicine)'

b. /au/
a.U 'to meet', u.TA.u 'to sing', hu.ru.MA.u 'to behave',
pa.U.ro 'St. Paul', do.NA.u 'River Donau (Danube)',
ra.ba.U.ru 'Rabaul', a.U.to 'out (in baseball)',
yun.gu.hu.RA.u 'Jung Frau', san.pa.U.ro 'Sao Paulo',
ro.na.U.do 'Ronald (football player's name)', sa.U.su 'south'

3 /ai/ fails to form a diphthong in this dialect when it is followed by a coda nasal. In this particular structure, /a/ forms one syllable, whereas /i/ constitutes an independent syllable with the following nasal, to avoid trimoraic syllables (Kubozono 2004): e.g., /de.ZA.in/ 'design', /sa.IN.kai/ 'signing party'. The same is true of /oi/ and /ui/: e.g., /ko.IN.syoo/ 'coin dealer', /ku.IN.bii/ 'queen bee'.

2.3 Diphthongs in Tokyo Japanese

As mentioned above, it is difficult to decide on the syllable structure of words in the mora-based system of Tokyo Japanese. However, it is not impossible to do so if we look carefully at phenomena that are sensitive to syllable boundaries. We will see two such phenomena here, one of which concerns the compound accent rule.

The compound accent rule of Tokyo Japanese places an accent on the final syllable of the first member if the second member is only one or two moras long (Akinaga 1981; McCawley 1968; Poser 1990). If the first member ends in a heavy, i.e., bimoraic, syllable, this rule places an accent on the head mora of this syllable, as indicated by an apostrophe (’) in (9). That is, a sudden pitch drop occurs between the two moras within a single syllable.

(9) a. oosaka + e’ki → oosaka’-eki ‘Osaka Station’
oosaka + si’ → oosaka’-si ‘Osaka City’

b. tookyoo + e’ki → tookyo’o-eki ‘Tokyo Station’
tookyoo + to’ → tookyo’o-to ‘Tokyo Metropolitan Government’

This accent test reveals contrasting behaviors of /ai/ and /au/ as shown in (10). In the examples in (10a), accent shift occurs within /ai/, suggesting that /ai/ forms a diphthong. In contrast, /u/ is accented in the examples involving /au/ as in (10b). This suggests that /au/ belongs to two separate syllables with a syllable boundary between the two vowels.

(10) a. hok.kai + do’o → hok.ka’i-doo ‘Hokkaido (place name)’
ma’.sai + zo’.ku → ma.sa’i-zo.ku ‘the tribe of Masais’

b. do’.na.u + ka.wa → do.na.u’-ga.wa, *do.na’u-ga.wa ‘The River Donau’
ri’n.dau + zin → rin.da.u’-zin, *rin.da’u-zin ‘the people of Lindau’

This analysis can be extended to other vowel sequences. We will consider here only those sequences that involve a falling or equal sonority. As illustrated in (11), /oi/ and /ui/ behave like a diphthong by attracting a compound accent on their initial vocalic element. In contrast, other vowel sequences behave like a heterosyllabic sequence, as shown in (12), with a compound accent on their final mora. For some speakers of Tokyo Japanese, the accent on /i/ might also be acceptable for words in (11), e.g., /rumoi’-si/ ‘Rumoi City’, but the accent on the first vowel of the vowel sequence is unacceptable for most speakers for the words in (12), e.g., */bide’o-situ/ ‘video room’. This difference is suggestive of the gap between /oi/-/ui/ and other vowel sequences.

(11) a. /oi/
ru'.moi + si' → ru.mo'i-si 'Rumoi City (place name)'
o.si.roi + ha.na' → o.si.ro'i-ba.na 'marvel-of-Peru (flower)'
to.ru.su'.toi + de'n → to.ru.su.to'i-den 'biography of Lev Tolstoy'

b. /ui/
sin.sui + si.ki' → sin.su'i-si.ki 'launching ceremony (of a ship)'
kai.sui + yo'.ku → kai.su'i-yo.ku 'swimming in the sea'
ko.tu'.zui + e'.ki → ko.tu.zu'i-e.ki 'marrow liquid'
ka'.mui + de'n → ka.mu'i-den 'Kamuiden (title of a cartoon)'

(12) a. /ao/
a.sa'.ga.o + i'.ti → a.sa.ga.o'-i.ti 'asagao (morning glory) market'
sa.o' + ta.ke → sa.o'-da.ke 'bamboo pole'

b. /ae/
ki.ga.e + si'.tu → ki.ga.e'-si.tu '(clothes) changing room'
na'.e + u.ri → na.e'-u.ri 'seedling peddler'

c. /eo/
bi'.de.o + si'.tu → bi.de.o'-si.tu 'video room'
bi'.de.o + ke'n → bi.de.o'-ken 'video coupon'

d. /oe/
a.ro.e + i'.ti → a.ro.e'-i.ti 'aloe market'

The idea that /ai/, /oi/ and /ui/ are the only real diphthongs in Tokyo Japanese can be further borne out by an analysis of baseball chants. According to Tanaka (1999), the phrase that baseball fans chant to cheer their favorite players consists of three musical notes plus a following pause ($). In the normal chant, this phrase corresponds to the player's name. Three-mora names show a straightforward pattern whereby each mora is associated with a musical note irrespective of the syllable structure involved: (13a) consists of three monomoraic syllables, (13b) is a bisyllabic name ending in a bimoraic syllable, and (13c) starts with a bimoraic syllable followed by a monomoraic one.

(13)	a. | ♩ ♩ ♩ $ |	b. | ♩ ♩ ♩ $ |	c. | ♩ ♩ ♩ $ |
	ma tu ki	sa ta n	sa n ta
	'Matsuki'	'Satan'	'Santa'

This mora-to-note correspondence is broken if the player's name consists of four or more moras. The basic rule in such cases is to assign the name's last syllable to the last musical note. Thus, the last note corresponds to the last syllable of the player's name, whether it is a heavy syllable as in (14a) or a light one as in (14b).

(14) a. | ♩ ♩ ♩ $ | | ♩ ♩ ♩ $ |
i ti roo da a win
'Ichiro' 'Darwin'

b. | ♩ ♩ ♩ $ | | ♩ ♩ ♩ $ | | ♩ ♩ ♩ $ |
na gasi ma zu ree ta sa nta na
'Nagashima' 'Zureta' 'Santana'

/ai/ and /au/ display different patterns with respect to this syllable-to-note association rule. Namely, /ai/ is assigned to the last musical note, but /au/ splits into two, with /a/ and /u/ being associated with different notes. This fact is illustrated in (15), where /o.ti.ai/ in (15a) patterns with /i.ti.roo/ in (14a), whereas /rin.da.u/ in (15b) behaves like /san.ta.na/ in (14b). This demonstrates that /au/ functions as a sequence of two syllables.

(15) a. |♩ ♩ ♩ $ | b. | ♩ ♩ ♩ $ |
o ti ai ri nda u
'Ochiai' 'Lindau'

This analysis, too, can be extended to vowel sequences other than /ai/ and /au/. Although people's names seldom end in a vowel sequence, we can readily use common nouns as test words instead. In (16)–(17), chunks corresponding to each musical note are separated by a dash /–/. This analysis shows that /makiroi/ and /oogui/ in (16) pattern with /i.ti.roo/ and /daa.win/ in (14a). This, in turn, suggests that /oi/ and /ui/ form one unit. On the other hand, /masanao/ in (17a) patterns with /na.ga.si.ma/ in (14b), suggesting that there is a syllable boundary between /a/ and /o/ in /ao/. Other vowel sequences in (17) also involve a syllable boundary.

(16) a. /oi/
ma – ki – roi 'McIlroy (proper name)'

b. /ui/
ki – n – sui 'Kinsui (proper name)'

(17) a. /ao/
ma – sa.na – o, *ma – sa – nao 'Masanao (proper name)'

b. /ae/
ma – tu.ma – e, *ma – tu – mae 'Matsumae (proper name)'

c. /eo/
i – wa.se – o, *i – wa – seo 'Iwaseo (proper name)'

d. /oe/
ka – wa.go – e, *ka – wa – goe 'Kawagoe (proper name)'

3 /ai/-/au/ asymmetry

We have seen in the preceding section that vowel sequences ending in /i/ readily form a diphthong, whereas those ending in other vowels do not. Most striking here is the asymmetry between /ai/ and /au/. Both consist of a low vowel followed by a high vowel, but they nevertheless exhibit contrasting patterns in syllabic organization. Interestingly, this asymmetry between /ai/ and /au/ can be observed in a wide range of phenomena in Japanese. The first person that noted this asymmetry is Motoko Katayama, who pointed out the following three facts (Katayama 1998). The first fact is that no SJ morphemes contain /au/, whereas a number of SJ morphemes contain /ai/ (section 3.1). Second, /au/ has shown a historical tendency to turn into the monophthong /oo/ in the adjectival morphology of native Japanese words, whereas /ai/ remains largely intact (section 3.2). And last, but not least, loanwords from English tend to retain the diphthong /ai/ in the English vowel sequence of /aiə/, while turning /au/ into /a/ in the sequence /auə/ (section 3.3 below).

In this section, we will consider these and other similar phenomena in detail. Section 3.1 discusses statistical frequencies with which the two vowel sequences occur in each of the three types of Japanese morphemes – SJ, native and foreign (see Kubozono, Introduction, this volume, and Nasu, this volume, for the lexical strata in Japanese). Section 3.2 considers the historical background of this synchronic state of affairs to understand why /ai/ enjoys a higher frequency than /au/ in the synchronic grammar. The next three sections (3.3 through 3.5) analyze the asymmetry between /ai/ and /au/ in the loanword phonology and morphophonology of contemporary Japanese.

3.1 Lexical strata and frequency

A noticeable difference between /ai/ and /au/ can be found in the frequencies with which the two vowel sequences occur in Japanese morphemes. In modern Tokyo Japanese, /ai/ occurs in a larger number of morphemes than /au/.[4] Of the three types of morphemes in Japanese, SJ morphemes exhibit this asymmetry in the most remarkable way. As Katayama (1998) pointed out, /ai/ is very commonly observed but /au/ is not attested at all in this type of morpheme. This has been borne out by Kubozono's (2005) analysis of all SJ morphemes listed in the appendix to a Japanese dictionary (Nagasawa 1959/1982). This analysis gives 407 SJ morphemes containing /ai/, but no instance containing /au/.

4 English shows a similar tendency, as noted in section 4.1.1 below. This may be linked to Maddieson's (1984) observation that the occurrence of /w/ usually implies the occurrence of /j/ in the same language.

A similar but more moderate asymmetry is observed in native Japanese (or so-called *Yamato*) morphemes. An analysis of native morphemes listed in the same appendix shows that /ai/ occurs in 63 morphemes, whereas /au/ is attested only in 29 morphemes (Kubozono 2005, 2008).[5] Most of the 29 native morphemes containing /au/ are verbal forms such as *au* 'to meet', *kau* 'to keep (an animal)' and *mau* 'to dance'. These forms may be analyzed as consisting of two morphemes rather than one. Thus, *kau* derives from the concatenation of a verbal stem /kaw/ and an ending /u/ just as *tobu* 'to fly' derives from /tob/ + /u/ (Vance 1987: 184), with the former but not the latter undergoing an independent process of /w/ deletion before a non-low vowel. Even if we assume that /au/ in these words belongs to one morpheme, there is phonological evidence, as we saw in section 2.3 above, which suggests that /au/ belongs to two syllables whereas /ai/ forms one syllable.

Finally, foreign morphemes seem to show a similar difference in frequency between /ai/ and /au/. It is certainly difficult to delimit an ever-increasing number of morphemes of this type in modern Japanese. However, the major source of foreign morphemes in modern Japanese is English (Sibata 1994; Irwin 2011),[6] where, as we will see in section 4.1 below, /ai/ occurs in a much larger number of words than /au/. Furthermore, there are several independent pieces of evidence that /au/, but not /ai/, tends to turn into a monophthong in a certain class of loanwords (sections 3.3–3.5 below). All these facts suggest that foreign morphemes, too, show an asymmetry between /ai/ and /au/, with the former appearing more frequently than the latter.

3.2 Vowel coalescence and historical stability

Given the remarkable difference between /ai/ and /au/ with respect to their frequencies in modern Japanese morphemes, a question that naturally arises is why such a difference is observed in the first place. This question can be answered in part by considering the history of the two vowel sequences in the language: see Takayama (this volume) for this and other sound changes.

The complete lack of /au/ in SJ morphemes may give the impression that it was absent in the inventory of vocalic phonemes in ancient or old Japanese.[7] This

5 /ai/ is generally more frequent than /au/ irrespective of the type of the preceding consonant. The only exception to this general tendency is the case where the word begins with a vowel sequence, i.e., with an onsetless syllable: seven native morphemes begin with /ai/, e.g., *ai* 'indigo (plant)', as opposed to nineteen native morphemes which begin with /au/, e.g., *au* 'to meet'.

6 According to Sibata (1994), 84% of loanwords used in contemporary Japanese are those that have been borrowed from English over the past one hundred years or so.

7 Following Kindaichi (1976), I assume five major periods in the history of Japanese: Ancient Japanese (up to Nara Period: –794), Old Japanese (Heian Period: 794–1191), Middle Japanese (Kamakura and Muromachi Periods: 1192–1603), Early Modern Japanese (Edo Period: 1603–1868) and Modern or present-day Japanese (1868–now).

impression turns out to be wrong, however. There is evidence that Japanese had this particular vowel sequence in at least some SJ morphemes (Kindaichi 1976). What happened then is that /au/ underwent a sound change called vowel coalescence, whereby it changed into a monophthong corresponding to /oo/ in Modern or earlier Japanese (see section 6 below for a general rule of vowel coalescence). This sound change took place at the end of Middle Japanese (in Muromachi Period) or earlier (Hashimoto 1950). The instances in (18) are cited from Kindaichi (1976: 159).

(18) au → oo 'cherry tree'
kau → koo 'high', 'fidelity'
kyau → kyoo 'capital', 'home town'

On the other hand, vowel coalescence did not occur obligatorily in morphemes containing /ai/. It did occur in casual speech at a later stage of Tokyo Japanese, where we now observe an alternation as shown in (19a) between careful and casual speech. This alternation is also observed in native Japanese words including those in (19b). However, this sound change did not occur in careful pronunciations in the dialect, nor did it penetrate into Kyoto Japanese and many other dialects. In fact, the monophthongal pronunciation for the original /ai/ is characteristic of casual speech in contemporary Tokyo Japanese.

(19) a. tai.gai ~ tee.gee 'usually'
sin.pai ~ sin.pee 'worry'
kyoo.dai ~ kyoo.dee 'brother'
dai.kon ~ dee.kon 'radish'
b. i.tai ~ i.tee 'painful, ouch'
hai.ru ~ hee.ru 'to enter'

A historical study of /ai/ and /au/ in native morphemes reveals a picture that is essentially identical to the one we saw above for SJ morphemes. As is well known, Japanese did not have any diphthong or any tautomorphemic vowel sequence at the beginning of its history. In the course of history, however, the language developed the two vowel sequences from /aCi/ and /aCu/ ("C" refers to any onset consonant) via consonant deletion processes called "i-onbin" and "u-onbin", respectively (Komatsu 1981).[8] The subsequent history of these newly created vowel sequences is almost parallel to that of /ai/ and /au/ in SJ morphemes. Namely, /au/ changed into a monophthong via vowel coalescence, whereas /ai/ remained intact except in very

8 Another source of vowel sequences is due to the deletion of a consonant in monomophemic words such as /kai/ (</ka.wi/) 'shellfish'.

colloquial (and often slangish) speech. Let us first consider the examples that Katayama (1998) gives for adjective+suffix sequences.[9]

(20) a. haya + i → hayai 'fast (conclusive form)'
haya + u → hayau → hayoo 'fast (adverbal form)'

b. taka + i → takai 'high, tall'
taka + u → takau → takoo

In sum, /ai/ has been quite stable in both native and SJ morphemes in the history of Japanese, whereas /au/ has undergone vowel coalescence into /oo/ in all SJ morphemes and most native morphemes.[10] Moreover, vowel coalescence affected /au/ at an earlier period of history than /ai/ in Japanese, where the coalescence of /ai/ into /ee/ remains an optional rather than obligatory phonological process.

Having looked at the striking difference between /ai/ and /au/ with respect to their historical stability, it is worth pointing out that this difference can be extended to other vowel sequences. Generally, vowel sequences whose second member is /i/ are more resistant than those ending in /u/ to the historical process of vowel coalescence. Thus, /oi/ and /ui/ show considerable stability and turn into a monophthong only in casual pronunciations of just adjectives in contemporary Japanese. On the other hand, their mirror-image counterparts, /eu/ and /iu/, almost obligatorily underwent coalescence. Some examples are given in (21).

(21) a. /oi/
Adj: sugoi ~ sugee 'great'
omosiroi ~ omosiree 'funny'
Noun: koi (no alternation) 'carp, love'
/ui/
Adj: atui ~ atii 'hot'
Noun: tuitati (no alternation) 'first day of the month'

b. /eu/ → /joo/
teuteu → tyootyoo 'butterfly'
neu → nyoo 'urine'
keu → kyoo 'today'
/iu/ → /juu/
iu → yuu 'to say'
riu → ryuu 'dragon'

9 Historically, all the words in the input arguably come from CVCVCV forms via consonant deletion: e.g., *hayasi* → *hayai*, *hayaku* → *hayau*.

10 Vowel coalescence did not occur in the final position of native verbs: e.g., /kau/, */koo/ 'to buy'.

Here, again, the obligatory coalescence processes in (21b) took place earlier than the optional processes in (21a) in the history of the language. The processes in (21b) occurred at the end of Middle Japanese (in Muromachi Period), according to Kindaichi (1976: 46), or earlier according to Hashimoto (1950: 89–90). The processes in (21a), in contrast, took place later in Early Modern Japanese (or in Edo Period).

While /oi/ and /ui/ developed in quite different ways from /eu/ and /iu/, /ei/ and /ou/ did not show such a noticeable difference. In fact, both /ei/ and /ou/ developed equally obligatorily into /ee/ and /oo/, respectively. These developments are illustrated in (22). However, Kindaichi (1976: 161) notes that these two developments, too, show a time difference, with /ou/ undergoing coalescence before /ei/ did (see also Takayama 1992).

(22) a. /ei/ → /ee/ <u>ei</u>yuu → <u>ee</u>yuu 'hero'
b. /ou/ → /oo/ <u>ou</u> → <u>oo</u> 'king'

In sum, vowel sequences ending in /i/ have been more or less stable in the history of Japanese, whereas those ending in /u/ have shown a striking tendency towards monophthongization. Moreover, vowel coalescence affected the former type of vowel sequences only after it affected the latter type in the course of the history. These historical facts seem responsible for the synchronic state of affairs discussed in the preceding section, and indeed reinforce our argument that /au/ is more marked than /ai/ in Japanese. Interestingly, Korean underwent a similar historical change, as we will see in section 4.2 below.

3.3 /aiə/ and /auə/ in loanwords

In addition to the two types of evidence we have so far seen, there are several other independent types of evidence for the relative markedness of /au/ over /ai/. All of these come from a phonological or morphological analysis of loanwords (see Kubozono Ch. 8, this volume, for more details about loanword phonology). Two of them concern the fate of English /ai/ and /au/ as they are borrowed into Japanese. Let us first consider the fact pointed out by Katayama (1998).

Katayama (1998) observes that /ai/ and /au/ before a schwa /ə/ are borrowed in different phonological forms in Japanese. They are illustrated in (23a,b), where the input forms show the source forms in English while the outputs show the loanword forms in Japanese.

(23) a. /aiə/ → /ai.ja(a)/
/taiə/→ /tai.ja/ 'tire'
/faiə/ → /fai.jaa/ 'fire'
/baiə/→ /bai.jaa/ 'buyer'

b. /auə/ → /a.waa/
/tauə/ → /ta.waa/ ‘tower’
/sauə/ → /sa.waa/ ‘sour’
/pauə/ → /pa.waa/ ‘power’
/auə/ → /a.waa/ ‘hour’
/flauə/ → /hu.ra.waa/ ‘flower’

The vowel sequence /aiə/ turns into a bisyllabic form /ai.ja/ with the palatal semivowel/glide /j/ added as the onset of the second syllable. This glide insertion is an independent process in Japanese that inserts /j/ in an onsetless syllable preceded by a non-back vowel: e.g., /pi.a.no/ → /pi.ja.no/ ‘piano’, /i.ta.ri.a/ → /i.ta.ri.ja/ ‘Italy’. On the other hand, the vowel sequence /auə/ undergoes the deletion of /u/ to yield the form /a.waa/. In this case, the labial glide /w/ is put before a schwa by an independent process that inserts a labial glide in an onsetless syllable preceded by a back vowel, e.g., /ko.a.ra/ → /ko.wa.ra/ ‘koala bear’ (Kubozono 2002a).[11] In this latter case, too, the resultant form is bisyllabic, with /w/ functioning as the onset of the second syllable. However, the two cases in (23) involve a crucial difference. In the case of /aiə/, both /a/ and /i/ survive in the resultant borrowed form, whereas /u/ is apparently lost in the case of /auə/. In loanwords, /au/ appears almost as freely as /ai/ in other phonological contexts, as exemplified in (24). However, it is clear that Japanese somehow avoids creating /au/ in the phonological context in (23). There is no comparable constraint on the occurrence of /ai/.

(24) /au.to/ ‘out’, /rau.do/ ‘loud’, /pau.daa/ ‘powder’

3.4 /ain/ and /aun/ in loanwords

Another piece of evidence suggesting the instability of /au/ as a syllable nucleus is found in the borrowing of /ain/ and /aun/ sequences. It is known that Japanese syllables are strongly constrained with respect to their maximal weight (Kubozono 1995, 1999a). In particular, they are subject to the general constraint prohibiting superheavy, i.e., trimoraic, syllables. This constraint, often called ‘trimoraic syllable ban’, applies specifically to long vowels and diphthongs as they appear with a coda consonant. If the original word contains a syllable consisting of a long (tense) vowel or diphthong plus a coda nasal as in *machine* and *ground*, this syllable is expected to yield a trimoraic syllable in Japanese with the nasal borrowed as a moraic coda nasal. This process is constrained by the syllable weight constraint, which forces trimoraic sequences into bimoraic ones. The most orthodox way to achieve this goal

11 Alternatively, the high back vowel /u/ may have turned into /w/. This does not affect the argument in this section, where the crucial difference between /aiə/ and /auə/ is at issue.

is to shorten the vocalic part, i.e., to shorten long vowels or to delete the second element of diphthongal vowel sequences. This shortening/deletion process, which Lovins (1975) described as "prenasal vowel shortening", is illustrated in (25). This process is equivalent to the well-known phenomenon of closed syllable vowel shortening in English and other languages by which long vowels with a coda consonant were historically shortened (Árnason 1980; Kubozono 1995).

(25) a. English /aun/ → Japanese /an/
gu.ran.do 'ground'
fan.dee.syon 'foundation'
me.rii.goo.ran.do 'merry-go-round'
wan.dan 'one-down' (in baseball)
tuu.dan 'two-down'
wan.ban 'one bound (ground ball)'

b. English /e:n/ → Japanese /en/
ren.zi 'range'
tyen.zi 'change'
a.ren.zi 'arrange'
su.ten.re.su 'stainless'
en.zye.ru 'angel'
ken.bu.rid.dzi 'Cambridge'

c. English /i:n/ → Japanese /in/
gu.rin.pii.su 'green peas'
ma.sin 'machine'
ku.in.bii 'queen bee'

d. English /o:n/ → Japanese /on/
kon.bii.hu 'corned beef'

The shortening process sketched in (25) is not a recent finding. Lovins (1975) described it many years ago and Kubozono (1994, 1995) proposed to explain it in terms of a constraint on the maximal weight of the syllable. However, these previous studies apparently overlooked an interesting asymmetry between /ain/ and /aun/. Namely, there seems to be no instance that involves shortening of /ain/ into /an/: that is, /ain/ is invariably manifested as such as shown in (26).[12]

12 This does not mean that /ain/ is accepted as a trimoraic syllable in Japanese. A careful analysis of the accentual behavior of /ain/ suggests that it actually consists of two syllables with a syllable boundary within /ai/, i.e., /a/ + /in/, in Tokyo Japanese (see Kubozono 1995, 1999a for details). This analysis can be further corroborated by evidence from Kagoshima Japanese, where /ain/ clearly splits into two syllables, /a/ + /in/ (Kubozono 2004). See Kubozono (Ch. 8, this volume) for details.

(26) sain 'sign', rain 'line, The Rhine', rain.ga.wa 'River Rhine', de.zain 'design', ko.kain 'cocaine'

This strongly contrasts with the fact that /aun/ is shortened to /an/ in many instances including those in (25a). There are exceptions to (25a), as we shall see shortly below, but this does not undervalue the contrastive behavior between /ain/ and /aun/. In fact, /au/ patterns with long vowels and tends to become a short monophthong. This means that the second element of /au/ behaves as if it were segmentally invisible when preceding a moraic nasal. This asymmetry between /ai/ and /au/ reinforces our argument that /au/, but not /ai/, is unstable in contemporary Japanese.

3.5 Stability in word formation

A fifth piece of evidence for the markedness of /au/ over /ai/ in Japanese stems from yet another fact showing the stability of /ain/ over /aun/. This evidence comes from a phonological analysis of the morphological process of compound truncation.

The most productive pattern of compound truncation in contemporary Japanese is to form a four-mora word by combining the initial two moras of one component word with those of the other (Ito 1990; Ito and Mester 1995; Kubozono 1999a, 2002b).[13] Some examples are given in (27), where 'L' and 'H' stand for light (monomoraic) and heavy (bimoraic) syllables, respectively, and { } denotes a foot boundary.

(27) a. LL+LL
se.ku.sya.ru ha.ra.su.men.to → {se.ku}{ha.ra} 'sexual harassment'

b. LL+H
po.ket.to mon.su.taa → {po.ke}{mon} 'Pokèmon, pocket monster'

c. H+LL
han.gaa su.to.rai.ki → {han}{su.to} 'hunger strike'

d. H+H
han.bun don.ta.ku → {han}{don} 'a half day off (= a half + holiday)'

As can be seen from (27), the truncation process in question is basically independent of syllable structure. That is, the utmost requirement is to yield a four-mora template – or, equivalently, a template consisting of two bimoraic feet. This default pattern, however, admits several types of exceptions, one of which concerns /aun/

13 Equally productive is the pattern whereby one component of a compound expression is entirely omitted with the other component remaining intact: e.g., *kontakuto renzu* → *kontakuto* 'contact lens', *keetai denwa* → *keetai* 'mobile phone', *suupaa maaketto* → *suupaa* 'supermarket'. See Ito and Mester (this volume) for the details about this and other word formation patterns.

sequences (Kubozono 2003b). As suggested above, there are quite a few exceptions to the shortening process in (25a). Some are given in (28), where syllable boundaries are not specified because of potential ambiguity.[14]

(28) saundo 'sound', maunten 'mountain', kaunsiru 'council', kaunto 'count'

These /aun/ sequences exhibit exceptional behavior in compound truncation. The rule sketched in (27) predicts that the words in (28) leave the initial two moras in this morphological process: e.g., /saundo/ → /sau/, /maunten/ → /mau/. However, what is actually observed is the pattern shown in (29), where the moraic nasal is retained while its preceding /u/ is deleted. This pattern is obtained whether /aun/ appears in the first component (29a) or in the second component (29b) (cf. Kuwamoto 1998b).[15]

(29) a. saundo torakku → {san}{tora} 'sound track'
b. buruu maunten → {buru}{man} 'Blue Mountain (coffee)'
buritissyu kaunsiru → {buri}{kan} 'British Council'
noo kaunto → {noo}{kan} 'no count (in baseball)'

While /aun/ exhibits an irregular pattern of truncation, /ain/ and /oin/ do not. There are not many truncated compounds that involve /ain/ or /oin/, but those that do actually follow the regular pattern by retaining the initial two moras of the trimoraic sequences, as exemplified in (30).

(30) a. donto maindo → {don}{mai}, *{don}{man} 'Don't mind'
b. zyointo bentyaa → {zyoi}{ben}, *{zyon}{ben} 'joint venture (business)'

Note here that the shortening of /au/ to /a/ in (29) is an entirely context-dependent phenomenon. /au/ follows the regular truncation pattern in (27) just as /ai/ does when it is not followed by a moraic nasal. As shown in (31), both /ai/ and /au/ retain their second mora when they appear before a syllable boundary.

(31) a. mai.ku.ro kon.pyuu.taa → {mai}{kon} 'micro computer'
poo.to ai.ran.do → {poo}{ai} 'Port Island (in Kobe)'
b. au.to do.rop.pu → {au}{do.ro} 'outdrop (in baseball)'

14 It is not clear yet what triggers pre-nasal shortening of /aun/ as in (25a) and what blocks this same process in the words in (28). It may be that pre-nasal shortening tends to affect /aun/ sequences in relatively long words and in old (as opposed to recent) borrowings.
15 Kuwamoto (1998b) makes the same observation but fails to notice that /aun/ behaves differently from /ain/.

In sum, the second mora of /aun/, i.e., /u/, is invisible to the morphological rule of compound truncation. Interestingly, long vowels and geminate obstruents often show a similar effect of invisibility in the same morphological process. This is illustrated in (32) and (33), respectively, where forms with an asterisk represent regular but unattested forms (Kubozono 1999a, 2002b, 2003a; Kuwamoto 1998a,b; Ito 2000).[16,17]

(32) a. paa.so.na.ru kon.pyuu.taa → {pa.so}{kon}, *{paa}{kon} 'personal computer'
suu.paa kon.pyuu.taa → {su.pa}{kon}, *{suu}{kon} 'super computer'
mee.ru to.mo.da.ti → {me.ru}{to.mo}, *{mee}{to.mo} 'e-mail friend'

b. daN.su paa.tii → {dan}pa, *{dan}{paa} 'dance party'
te.re.hon kaa.do → {te.re}ka, *{te.re}{kaa} 'phone card'

(33) a. bak.ku ten.kai → {baku}{ten}, *{bat}{ten} 'backward rotation (in gymnastics)'
a.me.ri.kan hut.to.boo.ru → {a.me}{hu.to}, *{a.me}{hut} 'American football'

b. po.te.to tip.pu.su → {po.te}ti, *{po.te}{tip} 'potato chips (fried potato)'

As mentioned in the preceding section, /au/ and long vowels show the same behavior in pre-nasal vowel shortening, i.e., they omit their second component. It is indeed interesting that /au/ patterns with long vowels rather than with /ai/ in compound truncation, too.

4 /ai/-/au/ asymmetry in other languages

In the preceding section, we have seen seven independent phenomena each of which shows an asymmetry between /ai/ and /au/ in Japanese. All these phenomena reveal the marked behavior of /au/ as opposed to /ai/: /au/ behaves as a very unstable vowel sequence and undergoes vowel coalescence or shortening (to /a/). In contrast,

---

16 It may be noticed that both long vowels and geminate obstruents exhibit different patterns of truncation depending on the location where they appear. Namely, when they appear in the medial position of truncated forms, a following mora tends to compensate for their shortening (32a)/(33a), whereas no such compensation occurs when they appear in the final position (32b)/(33b) (Kubozono 2002b, 2003b).

17 Long vowels and geminate obstruents do sometimes follow the regular truncation pattern: e.g., waa.do pu.ro.se.saa → {waa}{pu.ro} 'word processor', pa.to.roo.ru kaa → {pa.to}{kaa} 'patrol car = police car', a.ru.koo.ru tyuu.do.ku → {a.ru}{tyuu} 'alcoholism'; dan.zen top.pu → {dan}{to.tu} 'by far the best'.

/ai/ consistently functions as a diphthong in the language. All in all, /ai/ forms a much better syllable nucleus than /au/ in Japanese phonology.

Given this asymmetry between /ai/ and /au/, one naturally wonders why Japanese exhibits such an asymmetry at all. One way of tackling this question is to ask if the asymmetry is specifically observed in Japanese or if it is observed in a wide range of languages. Previous studies suggest the second possibility (Kubozono 2005, 2008).

4.1 English

There are at least two lines of evidence that show the marked behavior of /au/ as against /ai/ in English. They are both from Hammond's (1999) statistical work on the frequencies and phonotactics of English vowels in general.

4.1.1 Frequency

Hammond (1999) examined the frequencies of the fifteen monophthongs and diphthongs of English in a database of 20,000 words. This analysis has shown that /ai/ is far more frequent than /au/ irrespective of the length of words. The following table gives the number of each vowel in that database for words of different lengths (Hammond 1999: 106). Interestingly, the discrepancy between /ai/ and /au/ becomes larger as the word becomes longer. While it is unclear why /au/ is so rare in long words, the overall discrepancy between the two diphthongs is evident.

Table 1: Number of each vowel in words of different lengths

No. of syllables / Diphthong	1	2	3	4
/ai/	254	603	522	287
/au/	108	237	71	12

4.1.2 Phonotactic restrictions

Another interesting discrepancy is observed in the phonotactic restrictions imposed on the two diphthongs. This seems to account for the asymmetry in Table 1 at least in part. As noted by Hammond, /ai/ can stand before a larger number of consonants than /au/. Table 2 displays the strength of this cooccurrence restriction for the two diphthongs in word-final position: /–/ means the absence of an appropriate word.

Table 2: Cooccurrence restrictions between the diphthong and the following consonant

	/-p/	/-t/	/-k/	/-b/	/-d/	/-g/	/-f/	/-θ/
/ai/	ripe	right	like	bribe	ride	–	rife	blithe
/au/	–	bout	–	–	loud	–	–	mouth
	/-s/	/-ʃ/	/-v/	/-ð/	/-z/	/-ʒ/	/-tʃ/	/dʒ/
/ai/	rice	–	live	lithe	realize	–	–	oblige
/au/	mouse	–	–	mouthe	rouse	–	couch	gouge
	/-m/	/-n/	/-ŋ/	/l/	/r/			
/ai/	time	rine	–	rile	pyre			
/au/	–	town	–	cowl	hour			

As can be seen from this table, /ai/ can combine with coda consonants more freely than /au/: /ai/ combines with 16 out of 21 coda consonants, whereas /au/ combines with only 11 consonants. In fact, there are seven consonants that can stand after only one of the two diphthongs: six of them can follow /ai/, whereas just one, i.e., /-tʃ/, can follow /au/.

From a historical perspective, this is not an accidental asymmetry (Kubozono 2005). Modern English /ai/ and /au/ derive primarily from Middle English /i:/ and /u:/, respectively, which were diphthongized as part of the English Great Vowel Shift by about 1500 (Ekwall 1965/1975). However, diphthongization of ME /u:/ admitted a number of exceptions in the following phonological environments (34), whereas diphthongization of ME /i:/ admitted no such notable exceptions (Ekwall 1965/1975: 53).[18]

(34) a. before a labial: e.g., droop, room, tomb
b. before /k/: e.g., brook (verb)
c. before *r* + consonant: e.g., mourn, court, source
d. after /w/: e.g., wound, swoon, woo

(34a) probably accounts for the absence of /au/ before labial consonants in Table 2, i.e., before /p/, /b/, /f/, /v/ and /m/. Similarly, (34b) explains why /au/ + /k/ is not observed in the same table. The exceptions in (34a,b) can, in turn, be attributed to the phonetic fact that /u:/, but not /i:/, shares articulatory features with labial and velar consonants. In other words, the blocking of diphthongization of /au/ in (34a) and (34b) can be attributed to an assimilatory force to preserve the sequence of a vowel and a homorganic consonant. The same factor seems responsible for the blocking of /u:/ diphthongization in (34d), since /w/ also shares place

18 It seems that the Great Vowel Shift in German was not subject to this type of exception. In Modern German, /au/ occurs before labial consonants and /k/: e.g., *Raum* ‘room, space’, *Baum* ‘tree’, *Raub* /raup/ ‘robbery’, *Lauf* ‘run’, *tauglich* /tauk.liç/ ‘useful’.

features with /u/. In any case, it is clear that creation of /au/ was prohibited in certain phonological contexts, whereas creation of /ai/ was not subject to any such constraint in the history of English. In sum, /ai/ can cooccur with a coda consonant more freely than /au/ in English.

4.2 Korean

Korean shows a historical development of diphthongs in basically the same way as Japanese. Korean, in fact, seems a few centuries ahead of Japanese in historical development. According to Ahn and Iverson (2004), Middle Korean (15th century) had six diphthongs: /ɨi/, /ui/, /əi/, /oi/, /ai/ and /ɔi/. Interestingly, all of these end in /i/. Although it is not clear whether Old Korean had /au/, /eu/ and /iu/, it is interesting that tautosyllabic vowel sequences ending in /u/ were totally absent in the vowel inventory of Middle Korean. In other words, Middle Korean shows a striking asymmetry whereby all diphthongs end in /i/ and not in /u/. This is comparable to the situation of Early Modern Japanese, where every diphthong ended in /i/. In both languages, /ai/ and other vowel sequences ending in /i/ constitute very good syllable nuclei, whereas vowel sequences ending in /u/ such as /au/ and /iu/ do not.

Korean differs from Japanese in that it now has no diphthong. Of the six diphthongs permitted in Middle Korean, only /ɨi/ survived in the 18th–19th century Korean. This last diphthong has disappeared, too, with the result that no diphthong is present in the vowel inventory of present-day Korean. In contrast, modern Japanese still preserves /ai/, /oi/ and /ui/, which alternate with monophthongal forms in casual speech. In other words, monophthongization of /ai/, /oi/ and /ui/ was obligatory in Korean, while it remains an optional process in Japanese. In historical terms, this difference between Korean and Japanese can be interpreted as suggesting that the former is a few centuries ahead of the latter in the process of monophthongization. In synchronic terms, the same difference indicates that Korean is subject to a constraint prohibiting diphthongs in a more stringent way than Japanese.

4.3 Romanian

Another language that exhibits asymmetries between /ai/ and /au/ is Romanian (Kubozono 2005, 2008). Asymmetries in this language fall into two types. First, /au/ always constitutes two syllables in word-medial position, while /ai/ can be tautosyllabic in the same word position if /i/ is stressless. This is illustrated in (35). This asymmetry disappears in absolute word-final position, where both /ai/ and /au/ appear as a diphthong, as illustrated in (36).

(35) a. scaune /ská.u.ne/ 'chairs', caută /ká.u.tʌ/ 'they look for'
b. haine /hái.ne/ 'clothes', haită /hái.tʌ/ 'pack (of wolves)'

(36) a. au /áu/ 'they have', visau /ví.sau/ 'they were dreaming'
b. cai /kái/ 'horses', visai /ví.sai/ 'you (singular) were dreaming', malai /mʌ́.lai/ 'corn, maize', balai /bʌ́.lai/ 'blond'

A second type of asymmetry is observed when the vowel sequences in question occur before a word-final consonant (_C#). In this context, /au/ always splits into two syllables, while /ai/ and other vowel sequences ending in /i/ are accommodated within a single syllable. This is exemplified in (37a,b).

(37) a. flaut /flá.ut/ 'flute', balaur /ba.lá.ur/ 'dragon', faur /fá.ur/ 'craftsman', caut /ká.ut/ 'I look for'
b. raid /ráid/ 'raid', cuib /kúib/ 'nest', uit /úit/ 'I forget'

What (35) and (37) have in common is that /ai/ tends to constitute a syllable nucleus, whereas /au/ tends to refuse this integration. This is similar to the situation described in sections 2–3, where it was pointed out that /ai/ counts as one syllable and /au/ as two in Japanese. In both Romanian and Japanese, /ai/ constitutes a good syllable nucleus, while /au/ tends to form two nuclei, that is, two syllables.

5 Accounts for /ai/-/au/ asymmetry

We have seen in the preceding section that /ai/ and /au/ exhibit contrastive patterns across languages. In general, /au/ and other /Vu/ sequences are unstable as a diphthong, while /ai/ and other /Vi/ sequences form very stable syllable nuclei. While it is desirable to examine if this generalization holds in a greater number of languages, one important question to ask is why /au/ does not tend to constitute a good syllable nucleus. An optimal answer should also be able to explain why /au/ forms a natural class with /iu/, /eu/ and /ou/, and why the /ai/-/au/ asymmetry is observed in more than one language.

In the literature, Katayama (1998) proposes the constraint in (38) within the framework of Optimality Theory (Prince and Smolensky 2004).

(38) *[au]: [au] is not a good diphthong.

While very straightforward indeed, this account does not answer the questions raised above. It is no more than a restatement of the fact that /au/ is not a good

syllable nucleus. The fact that the asymmetry in question is observed across languages suggests that it should be attributed to some phonetic reason. To answer this question, Kubozono (2008) proposes a "better nucleus" hypothesis, which consists of the two principles in (39).

(39) Better nucleus hypothesis
V_1V_2 constitutes a better nucleus
(a) as V_1 is more sonorous than V_2.
(b) as the sonority distance is greater.

The notion "sonority" used here roughly corresponds to the acoustic parameter of F1, or to the articulatory parameter involved in amount of "openness" in the oral cavity – a more open jaw/lower tongue produces more sonorous sounds, e.g., vowels. According to Kasuya, Suzuki, and Kido (1968), the five vowels in Tokyo Japanese show the F1 values in Hz (and Bark) for male and female speakers as given in (40). From these values we can posit the sonority scale in (41).

(40)

	/a/	/o/	/e/	/u/	/i/
Female	888 (7.8)	483 (4.6)	483 (4.6)	375 (3.6)	325 (3.2)
Male	775 (7.0)	550 (5.2)	475 (4.5)	363 (3.5)	263 (2.6)

(41) Sonority scale: a > o, e > u > i > w > j

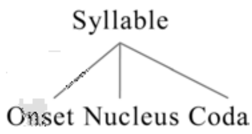
In light of this phonetic scale, /a/ and /u/ are more sonorous than /i/. Hence, /ai/ and /ui/ should form a better nucleus than /ia/ and /iu/, respectively, according to (39a). This prediction is fully supported by the fact that /ia/ and /iu/ do not generally form diphthongs: /ia/ often attracts a palatal glide via 'glide insertion', i.e., [ija] as in [pijano] for /pi.a.no/ 'piano', whereas /iu/ turns into a glide-vowel sequence, [ju], via 'glide formation' as in [karɯsjɯ:mɯ] for /ka.ru.si.u.mu/ 'calcium' (Kubozono Ch. 8, this volume). The principle in (39b) predicts, on the other hand, that /ai/ and /oi/ form better syllable nuclei than /au/ and /eu/, respectively. It also predicts that /ai/ is better than /ae/ as a syllable nucleus. These predictions all agree with the data given in the preceding sections. Specifically, the hypothesis in (39) accounts for the asymmetry between /ai/ and /au/, between /oi/ and /eu/ and between /ui/ and /iu/.

What must be noted here is that (39) is not just a hypothesis about syllable nucleus. Rather, it can be closely related to the generalization concerning syllable structure in language, generally known as "sonority sequencing principle" or "sonority sequencing generalization" (Sievers 1881; Blevins 1995).

(42) Sonority sequencing principle (Blevins 1995: 210)
Between any member of a syllable and the syllable peak, a sonority rise or plateau must occur.

This principle essentially defines the wellformedness of syllable structure with respect to a syllable constituent and the syllable peak (or nucleus). Given the simple syllable structure in (43), (42) entails that the syllable nucleus should be more sonorous than the onset and coda.

(43)



(39a) is a natural extension of the principle in (42): while (42) defines wellformedness with respect to the syllable nucleus and its neighboring elements, (39a) defines wellformedness with respect to the internal structure of the nucleus. In addition to this, (39b) recaptures structural wellformedness in a gradient rather than an absolute manner, by defining the "better-formedness" of the nucleus-internal structure.

This notion of relative optimality leads us to define the wellformedness of syllable structure in a more dynamic and principled way. First, it allows us to define the optimality of both onset and coda clusters as in (44).

(44) C_1C_2 forms a better onset/coda
a. as C_1 is less/more sonorous than C_2.
b. as the sonority distance is greater.

Again, (44a) is a simple extension of (42) by which the wellformedness of syllable structure is defined within the onset and the coda. (44b), on the other hand, defines the better-formedness of the onset and the coda on the basis of the notion 'sonority distance'. When combined, these principles account very nicely for what is actually observed in natural languages. First, (44a) explains the widely attested fact that /pl-/ forms a better onset cluster than /lp-/, as well as the fact that /-lp/ is a better coda than /-pl/. (44b), on the other hand, explains why /pl-/ is more likely to form an onset cluster than /pn-/, as well as why /-lp/ is more widely attested as a coda cluster than /-np/.[19]

The same idea of relative wellformedness can be applied to the relationship between the pre-nuclear glide and the nucleus of the syllable.[20] This possibility is summarized in (45).

19 This is in accordance with the sonority scale postulated by Ladefoged (1993: 246) for English.
20 It is assumed here that the pre-nuclear glide (such as /j/ as in /kjatto/ 'cat') is part of the onset rather than part of the nucleus (see note 1 above).

(45) Better glide-nucleus hypothesis
A glide and a vowel form a better Glide-Nucleus sequence
a. as the glide is less sonorous than the nucleus.
b. as the sonority distance between the glide and the nucleus is greater.

According to this hypothesis, /ju/ should be a better glide-nucleus sequence than /wi/ because the sonority distance between /j/ and /u/ is greater than that between /w/ and /i/. Similarly, /jo/ should form a better glide-nucleus sequence than /we/ since the sonority distance between /j/ and /o/ is greater than the distance between /w/ and /e/. These predictions must be tested by a wide range of empirical data, but as far as Japanese is concerned, they nicely agree with the phonotactic facts about the language. Namely, /ju/ and /jo/ are permitted, while /wi/ and /we/ are not in modern Japanese. The relevant phonotactic facts are given in (46) (see Pintér, this volume, for the emergence of new consonant-vowel sequences).

(46)	/ja/	*/ji/	/ju/	*/je/	/jo/
	/wa/	*/wi/	*/wu/	*/we/	*/wo/

Seen in this light, it can now be understood that the "better nucleus" hypothesis in (39) is only a subpart of a more general principle defining the wellformedness of syllable structure as a whole. The essence of this general principle is two-fold. First, sonority rises from the beginning of the syllable onset towards the peak of the syllable nucleus and falls from this peak towards the end of the syllable coda. In physiological terms, each syllable requires one jaw opening (e.g., the syllable nucleus) and each beginning and ending of the syllable requires a jaw closing (e.g., onset and coda). As one opens one's jaw (mouth) to make a syllable, the consonants should be ordered so that the most open consonants come closest to when the jaw is most open for the syllable nucleus.

Second, a better structure is obtained if the sonority distance between any adjacent elements is greater: within the onset, between the glide and the nucleus, within the nucleus, between the nucleus and the coda, and within the coda. This latter generalization can be summed up as follows.

(47) Sonority-distance principle
A given sequence of elements within the syllable, E_1E_2, is better-formed as the sonority distance between E_1 and E_2 becomes greater.

An advantage of this sonority-based account is that it defines the wellformedness of syllable structure in general, and not just the wellformedness of the syllable nuclei. Thus, one and the same account applies to (a) consonant clusters within the syllable onset, (b) those within the coda, (c) glide-nucleus sequences and (d) nucleus-coda sequences, as well as (e) two vocalic elements within the nucleus.

While the validity of the principle in (47) must be tested by a wide range of empirical data both from Japanese dialects and from other languages, it provides a principled account of the asymmetries between /Vi/ and /Vu/ in the syllable nucleus.

This said, it must be added that (47) cannot account for all the facts regarding the wellformedness of syllable nuclei. According to the sonority hierarchy in (41), the distance between /a/ and /u/ is greater than that of /u/ and /i/. The sonority-distance principle in (47) would then predict that /au/ should form a better nucleus than /ui/. Likewise, /au/ should be as good a nucleus as /oi/. These predictions are not borne out by the existing facts of Japanese, where /oi/ and /ui/ constitute better nuclei than /au/ both diachronically and synchronically. This suggests that it is necessary to look for an additional rule or principle in order to fully account for the data regarding the relative wellformedness of syllable nucleus.

6 Vowel coalescence

Before concluding this chapter, let us address the issue of vowel coalescence. We saw in the preceding sections that many vowel sequences, especially those that end in /u/, turned into a monophthong in the course of time. We also saw some vowel coalescence patterns that are productive in the synchronic grammar of contemporary Japanese. In this section, we will consider how many patterns there are in Japanese and how they can be generalized. (48) gives the twelve patterns that are listed in Kubozono (1999b), plus an additional pattern in (48m). They include sound changes that took place at a certain point of the history of the language as well as synchronic variations between careful and casual speech observed in present-day Japanese: /→/ indicates a sound change, whereas /~/ denotes a synchronic variation (vowel length in the output is ignored for the sake of simplicity).[21]

(48) a. /au/ → /o(o)/
haya + u → hayau → hayoo 'fast, quickly'
taka + u → takau → takoo 'high (adverb)'
kau → koo 'high', 'fidelity'
kyau → kyoo 'capital', 'home town'

b. /eu/ → /jo(o)/
teuteu → tyootyoo 'butterfly'
neu → nyoo 'urine'
keu → kyoo 'today'

21 Adjectives show the coalescence alternations in (48g,h) between careful and casual speech. They also admit a second casual speech form where the adjective stem undergoes final vowel lengthening: e.g., /atu-i/ → /a.tuu/ 'hot', /sugo-i/ → /su.goo/ 'great'.

c. /ou/ → /oo/
touzai → toozai ‘east and west’
sou → soo ‘priest’
you → yoo ‘errand’

d. /iu/ → /juu/
iu → yuu ‘to say’
riu → ryuu ‘dragon’

e. /ai/ ~ /e(e)/
naga-iki → nageki ‘long breath; grief’
itai ~ itee ‘Ouch!’
hairu ~ heeru ‘to enter’

f. /ei/ ~ /e(e)/
sensei ~ sensee ‘teacher’
de-iri ~ deeri ‘going in and out’

g. /ui/ ~ /i(i)/
atui ~ atii ‘hot’
samui ~ samii ‘cold’

h. /oi/ ~ /e(e)/
sugoi ~ sugee ‘great’
omosiroi ~ omosiree ‘funny’

i. /ae/ ~ /ee/
kaeru ~ keeru ‘to return’
kaeru ~ keeru ‘frog’

j. /oe/ ~ /ee/
tokoe ~ tokee ‘to the place’

k. /eo/ ~ /o/
mite-okoo ~ mitokoo ‘(Let me) look beforehand’
kangaete-oku ~ kangaetoku ‘(Let me) consider (it)’

l. /ea/ ~ /a(a)/
dewa ~ dyaa ‘well then’[22]
mite-ageru ~ mitageru ‘(I) will see it for you’
(hanako)-de-atta ~ (hanako)-datta ‘It was (Hanako)’

m. /oa/ ~ /a/
kono-aida ~ konaida ‘the other day’
tomo-are → tomare ‘in any case’

22 This analysis assumes that /dewa/ undergoes an independent process of /w/ deletion before undergoing vowel coalescence: i.e., /dewa/ → /dea/ → /djaa/.

Some of these patterns may be described as vowel deletion rather than coalescence since the output vowel is the same as one of the two vowels constituting the input: e.g., (48c–d, f–g, i–m). However, Kubozono (1999b) treats them in the same way as other patterns and proposes a unified rule for all the thirteen patterns in (48). The starting point of this proposal is to posit the following featural analysis of the five vowels in the language.[23]

(49) /i/ [+high, -low, -back]
/u/ [+high, -low, +back]
/e/ [-high, -low, -back]
/o/ [-high, -low, +back]
/a/ [-high, +low, +back]

Using these feature specifications, all the thirteen patterns in (48) can be accounted for by the rule in (50).

(50) $[\alpha \text{ high}, \delta \text{ low}, \varepsilon \text{ back}]_{V_1}$ $[\zeta \text{ high}, \beta \text{ low}, \gamma \text{ back}]_{V_2} \rightarrow [\alpha \text{ high}, \beta \text{ low}, \gamma \text{ back}]_{V3}$

This rule states that in vowel coalescence, the [high] feature of the output vowel (V_3) comes from the first element in the input (V_1), whereas the [low] and [back] features come from the second element (V_2). Under this analysis, the patterns in (48a), (48b), (48e) and (48h), for example, can be described as in (51)–(54), where underlined parts are the relevant features of the input that are inherited by the output.

(51) /au/ → /o/
$[\underline{\text{-high}}, \text{+low}, \text{+back}]_{V1}$ $[\text{+high}, \underline{\text{-low}, \text{+back}}]_{V2} \rightarrow [\text{-high}, \text{-low}, \text{+back}]_{V3}$

(52) /eu/ → /jo/
$[\underline{\text{-high}}, \text{-low}, \text{-back}]_{V1}$ $[\text{+high}, \underline{\text{-low}, \text{+back}}]_{V2} \rightarrow [\text{-high}, \text{-low}, \text{+back}]_{V3}$

(53) /ai/ → /e/
$[\underline{\text{-high}}, \text{+low}, \text{+back}]_{V1}$ $[\text{+high}, \underline{\text{-low}, \text{-back}}]_{V2} \rightarrow [\text{-high}, \text{-low}, \text{-back}]_{V3}$

(54) /oi/ → /e/
$[\underline{\text{-high}}, \text{-low}, \text{+back}]_{V1}$ $[\text{+high}, \underline{\text{-low}, \text{-back}}]_{V2} \rightarrow [\text{-high}, \text{-low}, \text{-back}]_{V3}$

23 /a/ is interpreted as a back vowel since it patterns with /u/ and /o/ in some phonological processes of Japanese such as vowel epenthesis in loanwords (see Kubozono Ch. 8, this volume). This interpretation does not bear upon the analysis of vowel coalescence proposed here, however, since the same analysis holds even if /a/ is interpreted as a non-back vowel.

This analysis is capable of accounting for all the thirteen patterns in a principled way. For one thing, it provides a unified account for not only those patterns that have been regarded as vowel coalescence but also those that have been characterized as vowel deletion, e.g., (48c, f–g, i–m). Second, it accounts for historical sound changes and synchronic variations in a unified manner. For example, (48e) represents a well-established case of historical change (/nageki/ 'grief') as well as a process that is in progress in modern Japanese (/itai/~/itee/ 'Ouch'). Third, it succeeds in generalizing three vowel coalescence patterns in the input: (i) vowel sequences involving a sonority fall (e.g., /ai/, /au/, /oi/), (ii) those where the two elements have an equal sonority (e.g., /iu/, /oe/), and (iii) those involving a sonority rise (/ea/, /oa/). Fourth, it can provide a unified account of coalescence in diphthongs, e.g., /ai/ → /e/, and coalescence as a resolution of vowel hiatus, or two adjacent vowels across a syllable boundary, e.g., /au/ → /o/ (see Kubozono Ch. 8, this volume, for other strategies to resolve hiatus in Japanese and other languages). And last, but not least, the rule in (50) can deal with tautomorphemic and heteromorphemic vowel sequences in the same way: e.g., the alternation between /sei/ and /see/ in (48f) occurs within a single morpheme, whereas that of /dei/ and /dee/ in (48f) involves a morpheme boundary.[24]

The generalization in (50) raises several new questions. For example, the palatal glide /j/ tends to appear in the output if the initial vowel (V_1) of the input sequence is a non-back vowel, e.g., (48b,d). However, this is not always the case, as in (48f,l). In (48k), the same input sequence results in the insertion of the glide in one example, i.e., /dewa/~/djaa/, but not in the other, i.e., /mite-ageru/~/mitageru/. It is interesting to ask why the latter case did not result in /mitjageru/ despite the fact this would be phonotactically perfect in the language. Overall, it is worth examining the mechanism whereby the glide is inserted in the output.

Equally interesting is the question of vowel length. The general tendency is to make the resultant monophthong long and thereby to preserve the phonological length of the input in the output. For example, /neu/ in (48b) is two moras long and so is its output form *nyoo* /njoo/. However, this is not always true, as can be seen from some of the examples in (48k,l). This raises a new question of when the mora length of the input is (not) preserved in the output or, equivalently, how the length of the output vowel is determined.

Furthermore, (50) itself does not predict the likelihood of vowel coalescence, i.e., which vowel sequence is more likely to undergo coalescence than others. We observe, for instance, some cases of /ae/ → /ee/ coalescence in (48i), but not any clear case of /ao/ → /oo/ coalescence: thus, /kao/ 'face' and /kaoru/ 'to smell' do

---

24 Another advantage of the generalization in (50) is that it correctly predicts that vowel coalescence does not occur in /ia/ or /ua/. The rule yields [+high, +low] in the output, which is a feature combination that cannot be phonetically interpreted. The fact that /ia/ and /ua/ do not show any pattern of coalescence supports the validity of the generalization in (50) indirectly.

not turn into /koo/ and /kooru/, respectively. In the preceding section, we considered phonetic reasons for the asymmetry between vowel sequences ending in /i/ and those ending in /u/, but it is not clear if they can provide a sufficient explanation for the difference between /ae/ and /ao/, or more generally, the likelihood with which a certain vowel sequence is subject to coalescence. This is another important question.

Finally, the cross-linguistic status of the rule in (50) is yet to be examined. Given the fact that vowel coalescence is a rather general and productive process in many languages of the world (see Casali 1996, 2011, among others), it is worth exploring whether the rule in (50) is specific to Japanese or applies to coalescence patterns across languages. This is another area where Japanese phonology may contribute to the general phonological theory.

7 Conclusion

In this chapter, we examined many phenomena regarding "diphthongs". We first looked at some phenomena showing that modern Tokyo Japanese permits only three diphthongs, /ai/, /oi/ and /ui/ (section 2). This analysis demonstrated an interesting asymmetry between vowel sequences ending in /i/ and those ending in /u/. We then focused on the asymmetry between /ai/ and /au/, and examined various facts that display the asymmetry (section 3). In descriptive terms, /ai/ tends to form a good syllable nucleus, whereas /au/ does not. In modern Japanese, /au/ is only found in foreign morphemes because it obligatorily underwent vowel coalescence or monophthongization in native and Sino-Japanese morphemes in the history of the language. Moreover, /au/ in foreign words is shortened to /a/ in some phonological contexts, e.g., before a schwa and before a coda nasal, in the process of borrowing. Furthermore, when it is segmentally realized in loanwords, /au/ is processed as two syllables, i.e., /a.u/, suggesting that syllabification is the last resort to resolve hiatus. In contrast, /ai/ functions as a good and stable diphthong throughout the history of the language.

The instability of /au/ as a syllable nucleus is not an isolated phenomenon in two ways. First, the asymmetry between /au/ and /ai/ is observed in other vowel sequences, too: vowel sequences ending in /u/, i.e., /eu/, /iu/ and /ou/, generally pattern with /au/, while vowel sequences ending in /i/, i.e., /oi/, /ui/ and /ei/, pattern with /ai/. Second, the instability of /au/ is observed across Japanese dialects as well as in some other languages (section 4). These facts suggest that the /ai/-/au/ asymmetry should be interpreted in a wider context.

In the second half of this chapter (section 5) we saw a phonetic account based on the notion "sonority". This account consists of two principles, one defining the overall shape of sonority contour within the syllable and the other defining the

sonority distance between adjacent elements within the same domain. Under the proposed analysis, /ai/ forms a better syllable nucleus than /au/ because of the greater sonority distance involved. The same account is responsible for the asymmetries between /oi/ and /eu/ and between /ei/ and /ou/, respectively. On the other hand, /iu/ is regarded as a bad syllable nucleus because it goes against the first principle: Sonority rises rather than falls between /i/ and /u/. In contrast, /ui/ forms a better nucleus than /iu/ because of the falling sonority involved.

In the final part of the chapter (section 6), we examined a variety of patterns underlying vowel coalescence in Japanese in search for a general rule that can account for the various patterns in a principled manner.

This chapter has raised as many questions as it has solved. We saw some of them in passing in the preceding sections. Aside from them, one important question for future work arises concerning the nature of the markedness of /au/ over /ai/. In the central part of this chapter, it was hinted that the /ai/-/au/ asymmetry is not restricted to Japanese phonology. We must pursue further cross-linguistic studies and confirm this point with many other languages.

It is also important to ask how the asymmetry between /ai/ and /au/ can be extended to other vowel sequences as well. It was hinted in passing that the relative markedness of /au/ over /ai/ may reflect a more general difference in markedness between vowel sequences ending in /u/, e.g., /iu/, /eu/, and those ending in /i/, e.g., /ui/, /oi/. This seems to hold at least in Japanese, but we should ask if the same is true of other languages.

If it turns out that /au/, /eu/ and /iu/ tend to form a less harmonic syllable nucleus than /ai/, /oi/ and /ui/ in many languages, we can then ask ourselves why that should be the case. We could tackle this question from various viewpoints, particularly from articulatory, acoustic and perceptual points of view. It will be interesting, for example, if we can experimentally show that /a/ and /u/ are perceptually more similar to each other than /a/ and /i/ are to each other, and therefore not easily tolerated in the same syllable, unless they coalesce into one vowel. This and other questions remain open for future work.

References

Ahn, Sang-Cheol and Gregory K. Iverson. 2004. On the diachronic inversion of Korean diphthongs. Paper presented at the 12th Manchester Phonology Meeting.

Akinaga, Kazue. 1981. Akusento shūtoku hōsoku [Laws of accent learning]. In the appendix to *Meikai nihongo akusento jiten* [Japanese Accent Dictionary]. Tokyo: Sanseido.

Árnason, Kristján. 1980. *Quantity in historical linguistics*. Cambridge: Cambridge University Press.

Blevins, Juliette. 1995. Syllable in phonological theory. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Oxford: Blackwell.

Casali, Roderic F. 1996. *Resolving hiatus*. Los Angeles: UCLA dissertation. (Garland Press, 1998).

Casali, Roderic F. 2011. Hiatus resolution. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1434–1460. Malden, MA and Oxford: Wiley-Blackwell.

Ekwall, Eilert. 1965. *Historische Neuenglische Laut- und Formenlehre*. Berlin: Walter de Gruyter. (English translation by Alan Ward. 1975. *A history of modern English sounds and morphology.* Oxford: Basil Blackwell.)

Hammond, Michael. 1999. *The phonology of English: A prosodic Optimality-Theoretic approach*. Oxford: Oxford University Press.

Hashimoto, Shinkichi. 1950. *Kokugo on'in no kenkyū* [Studies on Japanese phonology]. Tokyo: Iwanami.

Hirayama, Teruo. 1951. *Kyūshū hōgen onchō no kenkyū* [Studies on the tone of Kyushu dialects]. Tokyo: Gakkai no shishinsha.

Irwin, Mark. 2011. *Loanwords in Japanese* (Studies in Language Companion Series 125). Amsterdam/Philadelphia: John Benjamins.

Ito, Junko. 1990. Prosodic minimality in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.

Ito, Junko and Armin Mester. 1995. Binarity. Paper presented at the GLOW phonology workshop, Tromso, Norway, 3 June.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Ito, Emi. 2000. Gairaigo no fukugōgo ni okeru tokushu haku no arawarekata [Behavior of special moras in the truncation of loanword compounds]. *Nihon Joshidaigaku Kiyō* 49. 19–29. Nihon Women's University.

Kasuya, Hideki, Hisayoshi Suzuki and Kenichi Kido. 1968. Nenrei, seibetsu ni yoru nihongo 5 boin no picchi shūhasū to horumanto shūhasū no henka [Changes in fundamental frequencies and formant frequencies of the five Japanese vowels as a function of age and sex]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 24(6). 355–364.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology.* Santa Cruz, CA: University of California dissertation.

Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to the sounds of Japanese]. Tokyo: Ōfūsha.

Kibe, Nobuko. 2000. *Seinanbu Kyūshū nikei akusento no kenkyū* [Studies in the two-pattern accent systems in south-west Kyushu]. Tokyo: Bensei Shuppan.

Kindaichi, Kyosuke. 1976. *Nihongo no hensen* [Historical development of Japanese]. Tokyo: Kodansha.

Komatsu, Hideo. 1981. *Nihongo no on'in* [Japanese phonology]. Tokyo: Chūōkōronsha.

Kubozono, Haruo. 1994. Nihongo no onsetsu ryō ni tsuite [On syllable weight in Japanese]. *Kokugogaku* 178. 7–17.

Kubozono, Haruo. 1995. *Gokeisei to on'in kōzō* [Word formation and phonological structure]. Tokyo: Kurosio.

Kubozono, Haruo. 1999a. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo. 1999b. *Nihongo no onsei* [The sounds of Japanese]. Tokyo: Iwanami.

Kubozono, Haruo. 2001. Epenthetic vowels and accent in Japanese: Facts and paradoxes. In Jeroen van de Weijer and Tetsuo Nishihara (eds.), *Issues in Japanese phonology and morphology*, 113–142. Berlin: Mouton de Gruyter.

Kubozono, Haruo. 2002a. Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Journal of the Phonetic Society of Japan* 6(1). 79–97.

Kubozono, Haruo. 2002b. Temporal neutralization in Japanese. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology 7*, 171–201. Berlin: Mouton de Gruyter.

Kubozono, Haruo. 2003a. The syllable as a unit of prosodic organization in Japanese. In Caroline Fery and Ruben van der Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.

Kubozono, Haruo. 2003b. *Shingo wa kōshite tsukurareru* [How are new words coined?]. Tokyo: Iwanami.

Kubozono, Haruo. 2004. What does Kagoshima Japanese tell us about Japanese syllables? In Taro Kageyama and Hideki Kishimoto (eds.), *Nihongo no bunseki to gengo ruikei* [Analysis of Japanese and language typology], 75–92. Tokyo: Kurosio.

Kubozono, Haruo. 2005. [ai]-[au] asymmetry in English and Japanese. *English Linguistics* 22(1). 1–22.

Kubozono, Haruo. 2008. [ai]-[au] asymmetry: A phonetic account. In Haruo Kubozono (ed.), *Asymmetries in phonology: An East-Asian perspective*, 147–163. Tokyo: Kurosio.

Kubozono, Haruo. this volume. Introduction. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kuwamoto, Yuji. 1998a. Nihongo no ryakugo no on'in kōzō [The phonological structure of clipped words in Japanese]. *On'in Kenkyū* [Phonological Studies] 1. 161–168.

Kuwamoto, Yuji. 1998b. Ryakugo kara mita nihongo no chōjū onsetsu no kōzō ni tsuite [Truncation and the structure of trimoraic syllables in Japanese]. *Gengo Kagaku Ronshū* 2. 25–35. Tohoku University.

Ladefoged, Peter. 1993 [1975]. *A course in phonetics*, 3rd edition. Orland: Harcourt Brace & Company.

Lovins, Julia. 1975. *Loanwords and the phonological structure of Japanese*. Bloomington: Indiana University Linguistics Club.

Maddieson, Ian. 1984. *Patterns of sounds*. New York: Cambridge University Press.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCawley, James D. 1978. What is a tone language? In Victoria Fromkin (ed.), *Tone: linguistic survey*, 113–131. New York: Academic Press.

Nagasawa, Kikuya (ed.) 1959/1982. *Meikai kanwa jiten*. Tokyo: Sanseido.

Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Poser, William. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Saito, Yoshio. 1997. *Nihongo onseigaku nyūmon* [Introduction to Japanese phonetics]. Tokyo: Sanseido.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff and Richard T. Oehrle (eds.), *Language sound structure*, 107–136. Cambridge, MA: MIT Press.

Sibata, Takesi. 1994. Gairaigo ni okeru akusento kaku no ichi [The location of accent in loanwords]. In Kiyoji Sato (ed.), *Kokuo ronkyū 4: Gendaigo hōgen no kenkyū* [Studies in Japanese 4: Studies in modern Japanese and dialects], 338–418. Tokyo: Meiji Shoin.

Sievers, E. 1881. *Grundzüge der Phonetik*. Leipzig: Breitkoph and Hartel.

Takayama, Tomoaki. 1992. Nihongo ni okeru rensetsu boin no chōboinka [The changes of *Vu* hiatuses into long vowels in the history of Japanese language]. *Gengo Kenkyu* 101. 14–34.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Tanaka, Shin'ichi. 1999. Nihongo no onsetsu to yonhaku no tenpurēto: Senryū to puroyakyū seien ni okeru jiamari no bunseki [Syllables and four-mora templates in Japanese: Analysis of 'ji-amari' in *senryu* poems and baseball cheering]. In Spoken Language Working Group (ed.), *Bunpō to onsei II* [Speech and grammar 2], 261–290. Tokyo: Kurosio.
Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.
Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

III Morphophonology

Akio Nasu

6 The phonological lexicon and mimetic phonology

1 Introduction

The vocabulary of Japanese consists of different morpheme classes, each of which has a different etymological origin. As noted in a number of previous works (Martin 1952; McCawley 1968; Vance 1987; Shibatani 1990; Nishio 2002; Ito and Mester 1995a, 2003; among others), at least three classes have been distinguished traditionally: native, Sino-Japanese, and foreign. These three classes can be treated as two large groups as well. One is the native morpheme group, containing items which are indigenous to the language. Items in the native class are called *Wago* or *Yamato*.[1] The other group is the loanword class, which consists of Sino-Japanese and foreign vocabulary items. This type of vocabulary developed historically through the process of borrowing from other languages: the Sino-Japanese vocabulary is composed of roots that have been borrowed from Chinese, whereas the foreign vocabulary contains a large number of loans, most of which have been borrowed from European languages.[2]

In addition to the morpheme classes mentioned above, there is one more distinctive group of words in the Japanese vocabulary. This is the mimetic vocabulary, which consists of a rich variety of sound-symbolic items. Mimetic words express sounds of the external world in an imitative manner or symbolize states of objects, manners of movement, mental conditions, and so on. In this respect, mimetics can be treated as a special morpheme class which should be distinguished from the other morpheme classes. However, it must be noted that mimetic items are of native origin etymologically; most mimetic words have become established in the Japanese lexicon without any borrowing process. Due to their etymological status, the linguistic treatment of mimetic items has been controversial. On the one hand, some researchers regard mimetic words as a kind of native item. Labrune (2012: 13–14), for instance, remarks that mimetic words "belong to the Yamato class in the strict sense, even if they display a number of properties which may lead one to categorize them in a specific subclass."[3] On the other hand, in several theoretical accounts of the Japanese phonological lexicon, mimetics are treated as an independent lexical stratum from

1 See Irwin (2011: 5) for detailed comments on the terminology.

2 Based on a series of investigations (Ishiwata 1960; Umegaki 1963; Ueno 1980; Itō 2003; Hashimoto 2006; NLRI 1964, 1987), Irwin (2011: 25–26) presented data on the proportion of foreign words by donor language, and the dominant donor of present-day foreign words is English.

3 Irwin (2011: 6) also refers to this interpretation, noting that "…there is a strong claim for subsuming the mimetic stratum within native, as do many scholars who posit only three lexical strata."

Yamato (Ito and Mester 1995b; Fukazawa 1998; Fukazawa, Kitahara, and Ota 1998, 1999, 2002).

This chapter discusses the phonological properties of the Japanese lexicon with special attention to the peculiar relationship between Yamato and mimetics, in particular with respect to their phonological discrepancies and affinities. In the next section, we will make a general survey of morpheme classes in Japanese and of fundamental ideas about the phonological stratification of the lexicon. In section 3, we will examine a few stratum-specific phonological phenomena with a view to explicitly showing the phonological peculiarities of mimetics. In section 4, we will review the theoretical models presented in the previous literature on lexical stratification. Section 5 discusses the status of mimetics in the phonological lexicon of Japanese.

2 Stratification of the lexicon

2.1 Etymological classes and phonological properties

It is well known that morphemes belonging to distinct etymological classes exhibit different phonological properties from each other. Yamato items are distinctive in that they are subject to a well-known compound voicing process called *rendaku* (Martin 1952; Hamada 1960; Nakagawa 1966; Sakurai 1972; Kindaichi 1976; Vance 1987; Satō 1989; among others). In this process, the initial voiceless obstruent of the second element of a compound is voiced (e.g., *tori* 'bird' > *oya+dori* 'parent bird'). It occurs frequently in Yamato stratum, but loanwords do not undergo the process in general; see Vance (this volume) for a detailed discussion. In addition to rendaku, Yamato has a conspicuous restriction with respect to voicing; voiced obstruents (i.e., /b, d, z, g/) generally do not occur underlyingly in morpheme-initial position. Voiced obstruents are favored in morpheme-medial position in Yamato items such as *haba* 'width', *kuda* 'tube', *kaze* 'wind', *toge* 'thorn' (Hashimoto 1938; Komatsu 1981; NLRI 1984; Ito and Mester 1986; Labrune 2012; and see also Takayama, this volume, for a historical discussion).

Sino-Japanese items are characterized in terms of their unique shape. Sino-Japanese roots are in principle monosyllabic and variations of the syllable structure are limited to only the following four types: CV (*ka* 'course, department'), CVV (*doo* 'copper', *suu* 'number', *zei* 'tax', *kai* 'meeting', *rui* 'sort, class'), CVN (*kin* 'gold'), or CVCV (*betu* 'distinction, other', *koku* 'nation').[4] Though the last pattern has disyllabic structure, its underlying form is interpreted as monosyllabic /CVC/ and an epenthetic vowel (/u/ or /i/) is attached to the coda consonant to prevent a closed syllable from emerging (see Ito and Mester 1996 as well as Ito and Mester, this volume, for further

4 The onset consonant is optional in all four patterns. "N" denotes a moraic nasal that appears in the coda position.

discussion). In addition to monosyllabicity, palatalization of the onset consonant is characteristic of Sino-Japanese (Nakata 1982: 308–311). That is, Sino-Japanese contains a number of syllables in which the onset is palatalized, such as *tya* 'tea', *kyuu* 'emergency', and *myoo* 'strange, mystery'. McCawley (1968: 62–66) is an early theoretical attempt to account for the phonological diversity among morpheme classes with respect to the distribution of palatalized ("sharp" in his terminology) consonants.

The foreign stratum has many characteristic properties that distinguish the items involved from those in other strata. First, the emergence of voiced geminates, as in *beddo* 'bed', is a conspicuous feature of the foreign stratum (Martin 1952; Ito and Mester 1995a,b; Katayama 1998; Irwin 2011; Labrune 2012; among others). Second, the voiceless bilabial stop [p] can freely appear as a licit surface segment in foreign items such as *paipu* 'pipe', *puuru* 'pool', *supai* 'spy', etc. (McCawley 1968: 77–85; Shibatani 1990: 166–167; Ito and Mester 1995a,b; Labrune 2012: 70–77; among others). Third, foreign items frequently contain novel CV sequences which do not appear in Yamato and Sino-Japanese words (Hattori 1979; NLRI 1990; Ito and Mester 1995a,b; Katayama 1998; Irwin 2011; among others). For example, syllables containing the voiceless bilabial fricative [ɸ] can appear with no restrictions in foreign words such as *[ɸ]aito* 'fight', *[ɸ]iibaa* 'fever', *nai[ɸ]u* 'knife', *ka[ɸ]e* 'café', and *[ɸ]ooku* 'fork', whereas it can appear only before /u/ in Yamato and Sino-Japanese. NLRI (1990: 62–74) and Irwin (2011: 75) present the list of CV moras found only in foreign items; see also Pintér (this volume) and Kubozono (this volume).

2.2 Phonological stratification

As long as we restrict our attention to the facts mentioned above, each etymological class seems to have its own phonological properties that characterize that class exclusively. However, it is not always the case that a phonology-based characterization of lexical strata directly corresponds to the etymological classification of lexical items in a one-to-one fashion. Some phonological properties are shared among two or more morpheme classes. The following data, for example, show that items in etymologically different morpheme classes pattern together. In (1)–(2) and the rest of the chapter, dots (.) denote syllable boundaries.

(1) p~h alternation
  a. Yamato
    **po.si* (cf. *ho.si* 'star')
    **ya.pa.ri* (cf. *ya.ha.ri* 'likewise')
  b. Sino-Japanese
    **pak.ken* (cf. *hak.ken* 'discovery')
    **kai.pa.tu* (cf. *kai.ha.tu* 'development')

(2) Postnasal voiceless stop

a. Sino-Japanese		b. Foreign	
sinpo	'evolution'	*kyanpu*	'camp'
kantan	'easy'	*tento*	'tent'
sinsi	'gentleman'	*tyansu*	'chance'
denki	'electricity'	*tanku*	'tank'

The voiceless bilabial stop [p] cannot appear as a syllable onset following a vowel either in Yamato or in Sino-Japanese. In these two classes [p] is converted to [h], as exemplified in (1). With respect to the illegitimacy of [p], Yamato and Sino-Japanese pattern together. On the other hand, the data in (2) show that another grouping can be established, from which Yamato is excluded. An obstruent in postnasal position can be voiceless both in Sino-Japanese and in foreign items, while it must be voiced in Yamato forms such as *tonbo* 'dragonfly', *sinda* 'died', *kangae* 'thought, idea', etc. On the basis of the phonological patterns in (1) and (2), the following two groupings can be established for these morpheme classes.

(3) a. Yamato & Sino-Japanese // Foreign (singleton [p])
b. Yamato // Sino-Japanese & Foreign (postnasal voicing)

The phonology-based classification in (3) implies that the etymological partitioning of morpheme classes cannot account for all the properties of the synchronic configuration of the lexicon. The synchronic lexicon is, rather, organized gradiently, with some phonological properties overlapping between two (or more) morpheme classes.

With reference to the phonological classification of lexical items, it is notable that mimetic items exhibit quite distinctive behavior with respect to the phonological regularities discussed above. Although mimetic items are of native origin etymologically, they are exempt from the prohibition against singleton [p]; it appears as a licit segment, as exemplified in (4a). Moreover, as shown in (4b), mimetic items are subject to the process of postnasal voicing, just like Yamato items. ("-" denotes a morpheme boundary.)

(4) a. *pata-pata* 'pattering, flapping'
pika-pika 'shining, glittering'
pota-pota 'dripping'

b. *syonbori* 'dejectedly'
unzari 'disappointed'
zunguri 'stocky'

These data indicate the dual character of mimetics. While mimetics behave as if they belong to the native stratum with respect to the postnasal voicing (4b), they are the

opposite of Yamato with respect to the legitimacy of singleton [p] (4a). This complicated situation, summarized in (5), creates a problem concerning the lexical affiliation of mimetics, and it leads to the question of whether mimetic items constitute a separate class from Yamato or not.

(5)	a.	Yamato & Sino-Japanese // Foreign & Mimetics	(singleton [p])
	b.	Yamato & Mimetics // Sino-Japanese & Foreign	(postnasal voicing)

3 Phonological processes and lexical strata

As seen from the facts discussed so far, the phonological regularities observed in voicing and in the distribution of singleton [p] give rise to puzzles in the attempt to examine the phonological properties of the lexicon in Japanese. Thus, this section will review previous findings on regularities in voicing patterns and on the characteristic behavior of singleton [p] in Japanese phonology, with special attention to the relationship between Yamato and mimetics.

3.1 Voicing patterns and restrictions

Japanese has a voicing contrast in obstruents, and the distribution of voiced obstruents within morphemes is one of the features that differs among the lexical classes in the language. It has frequently been pointed out that the Yamato stratum has a restriction prohibiting underlying voiced obstruents morpheme-initially, in particular in Old Japanese (see Takayama, this volume, for details). Even in Modern Japanese, this restriction still exerts a profound effect on Yamato items; underlying voiced obstruents occur in principle only morpheme-internally in this stratum, as exemplified below.

(6)	*kubi*	'neck'	*hituzi*	'sheep'
	hada	'skin'	*kasegi*	'earning'
	suzu	'bell'	*kabuto*	'helmet'
	toge	'thorn'	*hadaka*	'naked'
	wasabi	'horseradish'	*kuzira*	'whale'
	karada	'body'	*kagami*	'mirror'

In contrast, the restriction does not hold at all in loanwords. Underlying voiced obstruents occur initially in a morpheme both in Sino-Japanese and in foreign words.

(7)	a.	Sino-Japanese		
		boo-koku	'national ruin'	(cf. *hoo-koku* 'report')
		den-ki	'electricity'	(cf. *ten-ki* 'weather')
		zin-tai	'human body'	(cf. *sin-tai* 'body')
		gin-ka	'silver coin'	(cf. *kin-ka* 'gold coin')

b. Foreign words

bakku	'back'	*buumeran*	'boomerang'
dansu	'dance'	*daiyamondo*	'diamond'
ziipu	'jeep'	*zeraniumu*	'geranium'
gaaru	'girl'	*guroobaru*	'global'

While root-initial voicing is prohibited in the Yamato stratum, it plays a significant role in indicating lexical contrasts among items in the loanword strata. In particular, as seen in (7a), the Sino-Japanese stratum contains a large number of minimal pairs distinguished only by voicing in the initial syllable of a root.

The distribution of voiced obstruents in the native stratum is subject to another characteristic regularity, namely, Lyman's Law.[5] It stipulates that a morpheme is permitted to have only one voiced obstruent. This restriction serves as a blocker of rendaku voicing in compounds.[6] If the second element of a compound already has a lexically specified voiced obstruent, rendaku cannot take place. Compare the Yamato compounds in (8a) and in (8b). While rendaku can take place in the former, it is blocked in the latter due to this restriction; see Vance (this volume) for a full analysis of Lyman's Law.

(8) a.

ike+bana	'flower arrangement'
yama+dera	'mountain temple'
hosi+zora	'starry sky'
iro+gami	'colored paper'
warai+banasi	'funny story'
dai+dokoro	'kitchen'
yo+zakura	'cherry blossoms at night'
go+gataki	'one's regular go partner'

b.

e+hude	'paintbrush'	(**e+bude*)
hitori+tabi	'solitary journey'	(**hitori+dabi*)
han+sode	'short-sleeved'	(**han+zode*)
tuti+kabe	'mud wall'	(**tuti+gabe*)
mato+hazure	'missing the mark'	(**mato+bazure*)
doku+tokage	'poisonous lizard'	(**doku+dokage*)
oo+sawagi	'spree'	(**oo+zawagi*)
onna+kotoba	'feminine speech'	(**onna+gotoba*)

5 As noted by Irwin (2011: 150), this regularity was originally stated by Motoori (1822), and it "was repeated by Lyman (1894) in English and thus became known as 'Lyman's Law'." Vance (1987: 136) also mentions this point.

6 Ito and Mester (2003: 34–36) demonstrate that Lyman's Law serves not only as a blocker of rendaku but also as a morpheme structure constraint in the Yamato stratum, indicating that simplex forms containing two voiced obstruents are systematically absent in the native vocabulary of Japanese. Thus, Lyman's Law holds in the Yamato stratum as an overall restriction prohibiting double obstruent voicing both in derived and non-derived environments.

In contrast, foreign words are exempt from the restriction of Lyman's Law. There are plenty of foreign words that contain two or more voiced obstruents. Ito and Mester (2003: 40–41) present numerous examples of Western loans disobeying the ban against double obstruent voicing. Some of these are shown below.

(9)	*baado*	'bird'	*kaadigan*	'cardigan'
	bondo	'bond'	*moogeezi*	'mortgage'
	daabii	'derby'	*zebura*	'zebra'
	gaaden	'garden'	*burudoozaa*	'bulldozer'
	gaido	'guide'	*dezitaru*	'digital'
	zyazu	'jazz'	*goburetto*	'goblet'
	binegaa	'vinegar'	*guroobaru*	'global'
	daburu	'double'	*riborubaa*	'revolver'
	guraidaa	'glider'	*sabuzyekuto*	'subject'

As for the Sino-Japanese stratum, the restriction is vacuously satisfied in all roots due to strict restrictions on the shape of morphemes. As discussed in Ito and Mester (1996, 2003) and Kurisu (2000), Sino-Japanese morphemes consist of at most two syllables, and only /t/ or /k/ can occupy the onset position of the second syllable (e.g., *betu* 'distinction, other', *koku* 'nation').

Based on the discussion so far, the phonological characteristics of each lexical stratum can be summarized as below.

Table 1: Phonological stratification (voicing regularity)

	Root-initial voicing	Double obstruent voicing
Yamato	NO	NO
Sino-Japanese	YES	–
Foreign	YES	YES

(YES = applies, NO = prohibited)

With respect to the distribution of voiced obstruents, the Yamato stratum is the most restricted and inflexible, while loanwords show no such restrictiveness.[7] Thus phonological stratification holds at least between Yamato and loanword items in terms of voicing properties.

7 Nishimura (2003, 2006) provides a few examples showing that Lyman's Law is in part active even in the foreign stratum. Foreign words frequently include voiced obstruent geminates, but they tend to undergo sporadic devoicing in certain types of loans. Nishimura observes that foreign words containing one or more voiced obstruent(s) in addition to the voiced geminate are more susceptible to sporadic geminate devoicing than words with only one (geminate) voiced obstruent. For instance, *baggu* 'bag' is more likely to be pronounced as *bakku*, but no such devoicing is expected in words such as *eggu* 'egg' (**ekku*).

3.2 Mimetic voicing

In addition to the strata discussed so far, mimetics are a distinct class with respect to the voicing restrictions. Some unique properties are observed in mimetic voicing as compared with other strata, in particular, with the Yamato stratum.

The first trait to be noted is that underlying root-initial voicing is not prohibited but rather favored in mimetics, although mimetic roots are of native origin etymologically. In mimetic vocabulary, root-initial voicing plays a significant role in sound-symbolic contrasts between items, and "[t]he contrast in voicing of initial obstruents is correlated with the semantic contrast of 'light/small/fine/thin' vs. 'heavy/large/coarse/thick'" as discussed by Hamano (1986: 106, 1998: 83). She also points out that "[f]or almost all C1C2 combinations in which C1 is a voiceless obstruent, there is a C1C2 counterpart in which C1 is voiced" in the mimetic vocabulary (Hamano 1998: 125).[8] Some representative examples of such word pairs, quoted from Hamano (1998: 125–126), are given below.

(10)	a.	*puwa-puwa*	light floating object
		buwa-buwa	large floating object
	b.	*tara-tara*	thick clear liquid
		dara-dara	thick murky liquid
	c.	*sawa-sawa*	the sound of a breeze
		zawa-zawa	the bustle of a crowd
	d.	*kata-kata*	clattering noise of a light object
		gata-gata	clattering noise of a heavy object

Root-initial voicing occurs productively, but voicing in root-medial position is comparatively less frequent in mimetics. According to the data presented by Hamano (1998: 41), which lists the types of consonants included in 366 bimoraic mimetic roots of the form C1V1C2V2, the number of items containing a voiced obstruent as C1 is 131, while the number of items containing a voiced obstruent as C2 is 54.[9] The following table shows the number of voiced/voiceless obstruents that occur as C1 or C2 in mimetic roots (based on Hamano's data).

8 "C1" and "C2" denote the consonants in the initial and the second syllable of a mimetic root.
9 Mimetic roots containing a voiced obstruent as C2 can be divided into four groups: (1) roots containing [b] as C2, (2) roots beginning with a sonorant (e.g., *mogu-* 'mumblingly'), (3) roots that are originally native morphemes but have been converted to adverbials with mimetic usage (e.g., *sizu-* 'calm, gently'), and (4) idiosyncratic exceptions (e.g., *kuda-* 'persistently'). Among these, roots of type (1) are a large majority, and they undergo systematic double-voicing as in *zuba-* 'boldly, frankly', as discussed in detail in section 5 (see also Hamano 1986, 1998, 2000 and Nasu 1999).

Table 2: Distribution of obstruents in disyllabic mimetic roots

C1 position				C2 position			
Voiced		Voiceless		Voiced		Voiceless	
b	41	p	44	b	28	p	4
d	19	t	26	d	5	t	67
z	23	s	28	z	14	s	40
g	48	k	36	g	7	k	72
		h	26			h	2
131 (45.0%)		160 (55.0%)		54 (22.6%)		185 (77.4%)	
291 (100.0%)				239 (100.0%)			

Root-initial position is disproportionately favored as the site for a voiced obstruent in mimetics, and interestingly enough, this pattern of distribution of voiced obstruents in roots is exactly the opposite of that in the Yamato stratum. While voiced obstruents in mimetics favor initial position, underlying voiced obstruents in Yamato roots appear almost exclusively in medial position, as already seen in (6).

The second point to be noted with respect to mimetic voicing is that mimetics obey Lyman's Law. The following data demonstrates that Lyman's Law is at work in the mimetic vocabulary. Mimetic roots containing two voiced obstruents (marked with "*") are ill-formed. (For the sake of simplicity, only bare root forms are shown.)

(11)	*beto-*	'sticky'	(**bedo-*)
	basa-	'with a rustle'	(**baza-*)
	dota-	'tramping'	(**doda-*)
	doka-	'bang, to dump'	(**doga-*)
	zito-	'damp'	(**zido-*)
	zuki-	'throbbing'	(**zugi-*)
	gasa-	'rustling'	(**gaza-*)
	gaku-	'wobbly'	(**gagu-*)

The third phenomenon related to voicing regularity is rendaku. Rendaku is known as a characteristic process of the Yamato stratum; on the other hand, it does not take place in mimetic words at all (Martin 1952: 49; Okumura 1955: 961–962; Sato 1989: 254). Reduplicated words are ideal examples to illustrate the opposite characteristics of Yamato and mimetic words (see also Vance, this volume, for discussion).

(12) a. Yamato

hito-bito	'people'
toki-doki	'sometimes'
saki-zaki	'the distant future'
kuni-guni	'countries'

b. Mimetics

pata-pata	'pattering'	(**pata-bata*)
toko-toko	'jog-trot'	(**toko-doko*)
saku-saku	'crunchy'	(**saku-zaku*)
kata-kata	'clattering'	(**kata-gata*)

Rendaku also does not occur in loanwords.[10] For example, the initial obstruent in the second member of a compound loanword such as *dezitaru+kamera* 'digital camera' is never altered to a voiced one: **dezitaru+gamera*. In this respect, mimetics are similar not to Yamato items but rather to loanwords.

In addition to the phenomena discussed so far, obstruent voicing in postnasal position must be taken into consideration. In the previous literature concerning phonological stratification of the lexicon, postnasal voicing has been frequently cited as one of the typical processes observed in Yamato phonology (Ito and Mester 1986, 1995a, 1995b, 2003; Ito, Mester, and Padgett 1995; Rice 1993, 1997, 2005; Ota 2004; among others). In Yamato words, postnasal voicing is observed both within morphemes (13a) and between morphemes (13b).

(13)	a.	*tonbo*	'dragonfly'	b.	/jom + te/	*yonde*	'reading'
		hotondo	'almost'		/kam + te/	*kande*	'biting'
		kangae	'thought'		/tob + te/	*tonde*	'jumping'
		tongaru	'pointed'		/sin + te/	*sinde*	'dying'

Forms which contain a voiceless obstruent immediately after a moraic nasal, e.g., **tonpo*, **hotonto*, **kankae*, and **tonkaru*, are ill-formed, in contrast to the existing forms in (13a).[11] The gerundive suffix *te* is subject to voicing and becomes *de* when it is attached to a verbal base ending with nasal or voiced bilabial segment, as shown in (13b); in this case, too, forms such as **yonte* are ill-formed. In contrast, postnasal obstruents in Sino-Japanese and foreign words do not undergo voicing. In these strata, voicing is contrastive in postnasal position, as shown below.

10 There are actually some exceptions in Sino-Japanese words such as *kabusiki+gaisya* 'limited liability company', *kuro+zatoo* 'brown sugar' etc. According to Sato (1989: 253), Sino-Japanese words which are used frequently in daily life tend to undergo rendaku. See Vance (this volume) for a full discussion.

11 There are in fact several counterexamples in which postnasal voicing fails to apply. Labrune (2012: 129), for example, points out some Yamato words such as *tanpopo* 'dandelion', *tinko* 'penis (child language)', *tinpira* 'young hooligan', etc., as exceptions of the process. (The initial consonant in *tinko* and *tinpira* are realized as an affricate [tʃ] phonetically.) The existence of these exceptional forms drives us to the question whether postnasal voicing is actually valid in the synchronic lexicon of the language. Ota (2004) examines counterexamples to postnasal voicing and argues that the learnability of lexical strata in Japanese is hard to explain given the synchronic distribution of postnasal voicing.

(14) a. Sino-Japanese

kon-ban	'tonight'	(cf. *kon-pan* 'this time')
kan-dan	'pleasant chat'	(cf. *kan-tan* 'easy')
kan-zen	'perfect'	(cf. *kan-sen* 'infection')
kan-gei	'welcome'	(cf. *kan-kei* 'relationship')

b. Foreign words

nanbaa	'number'	*ranpu*	'lamp'
torendo	'trend'	*tento*	'tent'
ziinzu	'jeans'	*sensu*	'sense'
tongu	'tongs'	*tanku*	'tank'

As for mimetics, there is reason to believe that postnasal voicing applies in this stratum. One word-formation pattern in mimetics produces emphatic forms in which a moraic consonant {C} is infixed as an intensifier (Hamano 1986: 137–139, 1998: 107–110). The intensifier appears as a voiceless obstruent when it is directly followed by a voiceless segment in a base, but it appears as a moraic nasal if it is directly followed by a voiced segment (Kuroda 1967; Hamano 1998: 35–36).

(15) a.

/sa-C-pari/	*sappari*	'clean, openhearted'
/ba-C-tari/	*battari*	'plump down, suddenly'
/ko-C-sori/	*kossori*	'secretly, stealthily'
/ga-C-kuri/	*gakkuri*	'in disappointment'

b.

/za-C-buri/	*zanburi*	'with a plop'
/ma-C-ziri/	*manziri*	'without a wink of sleep'
/ko-C-gari/	*kongari*	'perfectly fried'

The voiceless geminates in (15a) and the partial geminates in (15b) exhibit complementary distribution, and the latter contain a nasal + voiced obstruent consonant cluster. In contrast, a sequence in which a nasal is directly followed by a voiceless obstruent never appears in the emphatic form of a mimetic word, e.g., **sanpari*.

Given the discussion so far, the phonological properties involving voicing in the morpheme classes in Japanese can be summarized in Table 3, where the properties of the mimetic stratum are highlighted to make the point of interest clearer. Mimetics pattern together with Sino-Japanese and foreign items with respect to root-initial voicing and rendaku; a simple classification seems to hold, in which only the Yamato stratum is excluded from the lexical block composed of loanwords and mimetics. However, the matter is not so simple. Mimetics pattern together with Yamato items with respect to Lyman's Law and postnasal voicing, establishing another block which includes Yamato and mimetic items. Thus, we find the dual character of mimetics as indicated in the table.

Table 3: Lexical stratification on the basis of phonological regularities

	Yamato	Mimetics	Sino-Japanese	Foreign
Root-initial voicing	NO	YES	YES	YES
Rendaku	YES	NO	NO	NO
Lyman's Law	YES	YES	–	NO
Postnasal voicing	YES	YES	NO	NO

(YES = applies, NO = does not apply / prohibited)

3.3 Singleton [p] and p~h alternation

The voiceless bilabial stop [p] is an important segment for inquiries into the synchronic organization of the phonologically stratified lexicon in Japanese. Although it is a licit segment in the language as a whole, its distribution varies from stratum to stratum. To begin our discussion, let us observe the distribution of [p] in foreign words. It appears freely in this stratum, without any restrictions, as exemplified below.

(16) a. *pa.ri* 'Paris'
pin.ku 'pink'
poo.zu 'pause'
pa.ne.ru 'panel'

b. *su.pai* 'spy'
suu.paa 'supermarket'
kya.pu.ten 'captain'
he.ru.pu 'help'

c. *kyan.pu* 'camp'
su.ran.pu 'slump'
top.pu 'top'
su.top.pu 'stop'

[p] can appear either in word-initial position as an onset segment of a syllable (16a) or in word-medial position directly preceded by an open syllable (16b). We call [p] in these two environments "singleton [p]." [p] also appears directly after a closed syllable, as shown in (16c). In this case, [p] appears as the second half of a partial geminate [mp] or a total geminate [pp].

While [p] appears quite freely in the foreign stratum, its distribution is considerably restricted in the Yamato and Sino-Japanese strata. Singleton [p] cannot appear as a surface segment; it is subject to debuccalization, resulting in [h].

(17) a. Yamato
ha.na 'flower' (**pa.na*)
ya.ha.ri. 'likewise' (**ya.pa.ri*)

b. Sino-Japanese
hoo.ko.ku 'report' (**poo.ko.ku*)
ma.hoo 'magic' (**ma.poo*)

The *p~h* alternation (debuccalization) is a characteristic process in the Yamato and Sino-Japanese strata. McCawley (1968: 88) argues that [h] is excluded from the inventory of underlying segments both in Yamato items and in Sino-Japanese items and assumes that /p/ is the corresponding underlying form of the segment. He also argues that [h] is just a surface output form derived from /p/ by a series of feature changing rules (McCawley 1968: 124–125). On the other hand, in the foreign stratum, both /p/ and /h/ are interpreted as underlying segments (McCawley 1968: 88). These two segments actually show a phonological contrast in foreign items, as can be seen from the minimal pairs below.

(18)	a.	*patto*	'pat'	b.	*hatto*	'hat'
		pinto	'pint'		*hinto*	'hint'
		puragu	'plug'		*huragu*	'flag'
		pea	'pair'		*hea*	'hair'
		pooru	'pole'		*hooru*	'hall'

The evidence for the interpretation that [h] in Yamato is derived from underlying /p/ comes from the voicing alternation due to rendaku. In the general process of rendaku, the initial voiceless obstruent of the second member of a compound is voiced without any change in the place of articulation. That is to say, rendaku can be interpreted as a rule that changes the voicing specification and relates two obstruents which share the same place of articulation in underlying representation.

(19)	/watari+tori/	*watari+dori*	'migratory bird, migrants'	: /t/~/d/
	/tate+sima/	*tate+zima*	'vertical stripe'	: /s/~/z/
	/naki+koe/	*naki+goe*	'tearful voice'	: /k/~/g/

In line with this regularity, the underlying form of [h] should be /p/, since [p] has the same place specification as [b], which emerges as the output of rendaku voicing. This analysis is shown below. Underlying /p/ in a compound is voiced by rendaku, resulting in [b] in the output (20a); otherwise it surfaces as [h] due to debuccalization (20b).

(20)	a.	/nuri+pasi/	*nuri+basi*	'lacquered chopsticks'
	b.	/pasi/	*hasi*	'chopsticks'

As for Sino-Japanese, the pattern of gemination is one piece of evidence for regarding [h] as a surface segment derived from /p/. Sino-Japanese gemination takes place when a CVC-shaped root is directly followed by a root beginning with a voiceless obstruent.[12] In this process, regressive assimilation takes place at the morpheme

12 Stated more exactly, if the second C in the preceding CVC root is /t/, any voiceless obstruent in the onset of the following root triggers gemination, while only /k/ serves as a trigger of gemination if the preceding CVC root has /k/ as the second C. See Ito and Mester (1996) and Ito and Mester (this volume) for further discussion.

boundary; /t/ in the coda position of a preceding root is totally assimilated to the onset obstruent of the following root (Ito and Mester 1996).

(21) /bet-to/ *betto* 'special reserve (account)'
/bet-sei/ *bessei* 'specially made'
/bet-kan/ *bekkan* 'annex (to a building)'

If the onset segment in the second root is [h], the geminate cluster [pp] (not [hh]) appears. For example, the Sino-Japanese root compound *beppyoo* 'separate table' has the root *hyoo* 'table, list' as its second member. Thus, the underlying form of [h] should be interpreted as /p/; the underlying /p/ surfaces straightforwardly in the geminate (22a), but it is altered to [h] if it is not part of the geminate consonant (22b).[13]

(22) a. /bet-pjoo/ *beppyoo* 'annexed table'
b. /pjoo/ *hyoo* 'table, list'

3.4 [p] in mimetics

Mimetics are peculiar with respect to singleton [p]. In the mimetic stratum, singleton [p] can appear freely as a surface segment, as opposed to the Yamato and Sino-Japanese strata. As observed in the minimal triplet below, mimetic [p] has the status of a contrastive segment; it contrasts with both [b] and [h].

(23) a. *pura-pura* 'swinging'
b. *bura-bura* 'dangling'
c. *hura-hura* 'wobbling'

Singleton [p] in mimetics is peculiar not only in its contrastive status but also in its great lexical frequency. According to Hamano (1986, 1998, 2000), [p] is frequent in the initial position (C1) of bimoraic mimetic roots. The lexical frequencies of consonants in C1 position in mimetic roots are shown in the Table 4 below, which is based on Hamano's (1998: 41) data.

Table 4: Lexical frequency of consonants in C1 position of reduplicative mimetics

p	44	*d*	19	*y*	6
b	41	*s*	28	*k*	36
h	26	*z*	23	*g*	48
m	24	*n*	18	*w*	4
t	26	*r*	0		

13 [p] also appears as the second half of the partial geminate [mp] in Sino-Japanese root compounds such as *sinpai* 'anxiety'. See Ito and Mester (1996) for details.

Although the most frequent segment in the C1 of reduplicative mimetics is *g*, if we limit our observations to voiceless obstruents, we find that *p* is the most frequent segment among them (*p*: 44 > *k*: 36 > *s*: 28 > *t*: 26 = *h*: 26). In contrast, the lexical frequency of *p* in Yamato morphemes is quite low. The data provided by NLRI (1984) show this point clearly. The following table, extracted from NLRI (1984: 25), shows the lexical frequency of consonants appearing in the initial position of bimoraic Yamato nouns.[14]

Table 5: Lexical frequency of consonants in C1 position of bimoraic Yamato nouns

k	153	*d*	25	*m*	124
g	20	*n*	94	*r*	6
s	128	*h*	132	*w*	20
z	15	*b*	36	*N*	0
t	125	*p*	2		

It is noteworthy that Yamato and mimetic items are diametrically opposed with respect to the frequency of singleton [p], even though they are all indigenous to the Japanese language in terms of their etymological origin.

Let us now summarize the main points that have been made in this section. With respect to the legitimacy of singleton [p], the phonological properties of each morpheme class and the relationships among classes can be illustrated as below.

Table 6: Lexical stratification on the basis of singleton [p]

	Yamato	Mimetics	Sino-Japanese	Foreign
singleton [p]	NO	YES	NO	YES

(YES = allowed, NO = disallowed)

In spite of being etymologically akin to Yamato items, mimetics are exactly opposite in terms of the legitimacy and lexical frequency of singleton [p]. Mimetics pattern together with foreign vocabulary in that [p] appears freely as a contrastive surface segment. The Yamato stratum, on the other hand, behaves like the Sino-Japanese stratum, which is clearly distinct etymologically. This kind of discrepancy between phonology and etymology is not surprising, but we have to note that the inconsistent relationships among lexical classes observed with respect to singleton [p] are somewhat different from what we observed in voicing patterns. In the case of voicing, as shown in Table 3, the Yamato and Sino-Japanese strata do not pattern together with respect to any of the criteria. When it comes to singleton [p], however, they behave identically. This kind of complicated relationship among morpheme classes has been at the core of many discussions concerning the phonological stratification of the lexicon in Japanese.

14 "N" indicates moraic nasal. As for *y*, it is categorized as part of a *yō-on* (a syllable containing a palatalized onset) and segregated from the plain consonant category in NLRI's (1984) data.

4 Theoretical accounts of phonological lexicon

In this section, we will review theoretical treatments of the phonological lexicon in Japanese, taking up three representative proposals developed in the framework of Optimality Theory.

4.1 Core-periphery model

One possibility for a formal account of the stratified nature of the lexicon is to assume totally partitioned sublexicons for each etymological class in the language. In this approach, each separate sublexicon involves stratum-specific phonological rules identified by morpheme-class features such as [+Yamato] or [–Foreign], as proposed in McCawley (1968). The sublexicon model may appear at first to be feasible, but we cannot overlook the fact that some phonological properties do not display the simple distributions in the lexicon that this model leads one to expect. Some properties are shared by more than one morpheme class, as already discussed, and the boundaries implied by the phonological regularities do not correspond exactly to the domains of each sublexicon. A sublexicon model that postulates totally partitioned mini-lexicons does not have an effective way of dealing with such a complicated inter-stratum relationship.

Ito and Mester (1995a) propose a novel model of the synchronic lexicon in which no partitioned sublexicons are postulated. Instead, one unitary abstract domain is assumed to represent the notion of the lexicon as a whole. This abstract domain involves several smaller domains, each of which is defined by a markedness constraint, and lexical items within a certain domain are required to pattern together with regard to the regularity imposed by that markedness constraint. Ito and Mester (1995a) demonstrate that this notion of “constraint domain” provides an explicit account of the synchronic configuration of the phonological lexicon. Moreover, what has primary significance in their model is the notion of a “core-periphery” relationship among lexical strata. The configuration of a lexicon with this core-periphery structure can be illustrated as below.

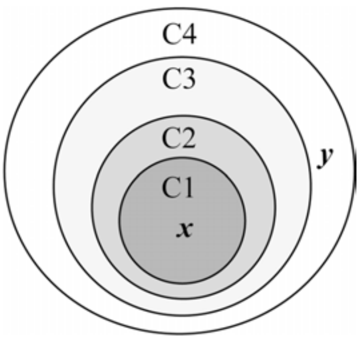


Figure 1: Core-periphery configuration of a lexicon

The implicational relationship represented by the concentric circles in the figure is equivalent to a diagram that exhibits the degree of nativization. The area located closest to the core of the concentric circles corresponds to the space of native (or fully nativized) items, whereas the space on the periphery contains less nativized lexical items.

Another significant insight provided by the model is that the relative priority of each phonological constraint in the lexicon can be captured by an implicational relationship among the constraint domains. Items contained in the core area, such as native morphemes, are subject to many more phonological restrictions than items in the peripheral area. For instance, items indicated by "*x*" in the Figure 1 must obey all of the constraints (C1–C4), while those represented by "*y*" are exempt from the restrictions imposed by the constraints C1, C2, and C3. That is to say, "As the periphery is approached, many of the constraints cease to hold, or are weakened in systematic ways," as explained in Ito and Mester (1995a: 824).

The core-periphery model has an advantage over the sublexicon model, since it provides a mechanism that yields different phonological patterns within a single stratum. To show how well the model works, Ito and Mester (1995a) conduct a case study examining the distribution of plain and palatal moras in foreign words. The Japanese coronal stops /t, d/ ordinarily change into the alveopalatal affricates [tʃ, dʒ] when they are directly followed by the high front vowel /i/. CV-moras without this palatalization, i.e., [ti] and [di], never appear in the native, Sino-Japanese, or mimetic strata. In the foreign stratum, however, items are divided into two groups in this respect. One is the group in which palatalization takes place just as in the native stratum, and the other is the group without palatalization. While the pattern in (24a) is observed in relatively assimilated items, forms like those in (24b) are common among recent unassimilated loans (examples from Ito and Mester 1995a: 828).

(24)	a.	*[tʃ]iimu*	'team'	b.	*[t]iin*	'teen (ager)'
		[tʃ]iketto	'ticket'		*paa[t]ii*	'party'
		[dʒ]irenma	'dilemma'		*[d]isuku-zyokkii*	'disc jockey'

The lexical status of these two patterns can be accounted for by means of a constraint domain delimited by a sequential constraint *TI, which prohibits nonpalatal coronal consonants followed by /i/. As illustrated below, the forms in (24a) are included in this domain, while those in (24b) are outside it.

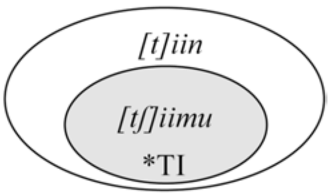


Figure 2: Lexical stratification by means of constraint domains

Note that the domain labeled by *TI occupies the inner of the two concentric circles, and that the lexical status of the example words differs with respect to the notion of distance from the lexical core. A location closer to the core indicates that *[tʃ]iimu* 'team' is more assimilated, conforming to a conservative pattern, while *[t]iin* 'teen (ager)' is more innovative, as reflected by its greater distance from the core.[15] Even though these two words have an identical etymological origin as "foreign" loans, their actual status in the lexicon differs in terms of the degree of nativization. Thus the core-periphery model, taking advantage of the notion of a constraint domain, can provide a realistic picture of the organization of the synchronic lexicon without any reliance on the etymological origin of lexical items.

4.2 Lexical stratification via constraint reranking

The internal stratification of the synchronic lexicon can also be accounted for in terms of an optimality-theoretic device in which the notion of constraint ranking plays a key role. Ito and Mester (1995b) suggest that the core-periphery organization of the phonological lexicon can be captured by means of a constraint hierarchy consisting of a few markedness constraints. The fundamental insight of this idea is that elements in the lexical core and those in the peripheral area have different characteristics with regard to the degree of satisfaction of the constraints. While the former must fulfill all of the markedness constraints, the latter, which are outside the lexical core, can violate most of the constraints in the hierarchy. Ito and Mester (1995b: 183) present the following three points as the central results of their study. First, the lexicon of a language as a whole is governed by a single, invariant ranking of markedness constraints. Second, lexical stratification is explained by the mechanism of constraint "reranking". And third, reranking is limited to faithfulness constraints.

Ito and Mester (1995b) propose the four syllable-related markedness constraints in (25). These constraints are in an implicational relationship with each other, with the ranking shown in (26). (In the literature on constraint-based accounts of lexical stratification, NoVoiGem, No-[P], and PostNasVoi are frequently indicated as *DD, *P, and *NT, respectively. We will follow this practice hereafter.)

(25) a. SyllStruc
Constraints defining the basic syllable canon of Japanese, including NoComplexOnset, NoComplexCoda, CodaCond.

15 Needless to say, this is a somewhat idealized analysis, abstracting away from the actual behavior of individual foreign items. Irwin (2011: 82–83) points out that an "intermediate layer" is observed in actual usage of foreign items, in which a large number of "conservative and contemporary doublets" are found.

b. NoVoiGem (*DD)
Geminate obstruents must be voiceless. (No voiced obstruent geminates.)

c. No-[P] (*P)
[p] is licit in doubly linked configurations. (No singleton [p].)

d. PostNasVoi (*NT)
Post-nasal obstruents must be voiced. (No voiceless postnasal obstruents.)

(26) SyllStruc » NoVoiGem (*DD) » No-[p] (*P) » PostNasVoi (*NT)

Lexical items in a stratum which is subject to a lower ranked constraint must obey all of the higher ranked constraints, but not vice versa. That is, the ranking in (26) corresponds to the inclusion organization illustrated in the Figure 3, which is composed of a small number of constraint domains labeled with the constraints introduced in (25). For instance, items in the lexical stratum in the innermost domain (typically Yamato items), are subject not only to *NT but also to the other constraints in the outer domains. In contrast, items in the outermost domain (typically unassimilated loans) are exempt from the constraints other than SyllStruc.

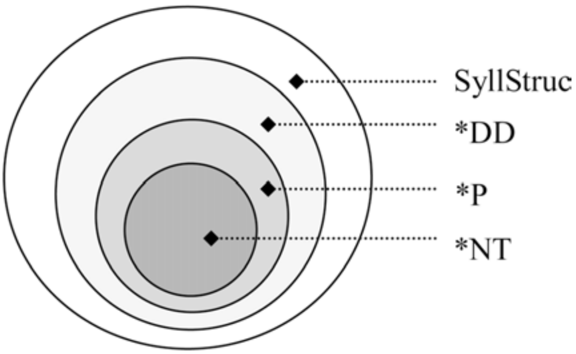


Figure 3: Constraint ranking and core-periphery organization

The implicational organization illustrated in Figure 3 represents a static map of the distribution of constraint domains, but the map itself cannot directly account for the gradual character of the stratified lexicon. In addition to the markedness constraints, faithfulness constraints and their place in the constraint hierarchy play a crucial role. As frequently discussed in connection with the classical model of Optimality Theory (Prince and Smolensky 1993/2004), the scope of application of markedness constraints is determined through inevitable conflict with faithfulness constraints. In the invariant ranking of the markedness constraints in (26), there are

four sites where faithfulness constraints can intervene, as illustrated in (27). (The family of faithfulness constraints is treated as a single unit and designated as FAITH in the model.)

(27) Constraint rankings via reranking of FAITH

<table>
<tr><th>a. Yamato</th><th>b. Sino-Japanese</th><th>c. Foreign</th><th>d. Alien[16]</th></tr>
<tr>
<td>SyllStruc
|
*DD
|
*P
|
*NT
|
FAITH</td>
<td>SyllStruc
|
*DD
|
*P
|
FAITH
|
- - - - - - - - - - - -
*NT</td>
<td>SyllStruc
|
*DD
|
FAITH
|
- - - - - - - -
*P
|
*NT</td>
<td>SyllStruc
|
FAITH
|
- - - - - - - -
*DD
|
*P
|
*NT</td>
</tr>
</table>

The ranking in (27a) represents the grammar observed in the core area of the phonological lexicon. In this ranking, all of the markedness constraints must be satisfied, even at the cost of violating FAITH. On the other hand, in each of the other strata, the markedness constraints lose their power beneath the dotted line. In this respect, we can say that FAITH serves as a kind of switch that makes the markedness constraints inert when they are dominated. The stratified nature of the phonological lexicon is explained by means of this machinery of constraint *reranking*. For instance, lexical items are exempt from *NT in the ranking (27b) due to the one-step promotion of FAITH as compared with the ranking in (27a). The rankings in (27c) and (27d) emerge as a result of further reranking of FAITH, which necessarily entails the weakening of *P and *DD. That is to say, promotion of FAITH in the ranking means that the phonological grammar comes closer to the periphery, where phonological patterns unique to unassimilated loans emerge.

4.3 Multiplication of faithfulness constraints

The reranking model (27) is significant in that it models the gradual nature of the stratified lexicon making use of a fundamental concept of Optimality Theory. However, Fukazawa (1998) and Fukazawa, Kitahara, and Ota (1998) point out that there is an empirical problem for the reranking model. Although the reranking

16 "Alien" refers to a class of unassimilated foreign items, many of which have been borrowed into Japanese lexicon recently.

model successfully captures the stratum-specific characteristics observed in each lexical stratum, it cannot account for the phonological properties of hybrid compounds, which are composed of morphemes from different lexical strata. For example, in the hybrid compound *tonbo-kenkyuuka* 'dragonfly-researcher', in which the first member is a Yamato morpheme and the second member is a Sino-Japanese word, the markedness constraint *NT must be satisfied in the former, while it is violated in the latter. In the reranking model, since each lexical stratum requires a different niche for FAITH in a single constraint hierarchy, the following ranking paradox inevitably arises. In order to obtain the correct output for the first element *tonbo*, *NT must be ranked above FAITH, while it must be dominated by FAITH with respect to the Sino-Japanese second element *kenkyuuka*. The reranking model cannot explain the phonological structure of the hybrid, as shown below ("☹" denotes a wrong output).

(28) a. *NT » FAITH

/tonpo-kenkyuuka/		*NT	FAITH
i.	to**np**o-ke**nk**yuuka	*!*	
☹ ii.	to**nb**o-ke**ng**yuuka		**
iii.	to**nb**o-ke**nk**yuuka	*!	*

b. FAITH » *NT

/tonpo-kenkyuuka/		FAITH	*NT
☹ i.	to**np**o-ke**nk**yuuka		**
ii.	to**nb**o-ke**ng**yuuka	*!*	
iii.	to**nb**o-ke**nk**yuuka	*!	*

Although the desired winner is *tonbo-kenkyuuka* in (28-iii), neither ranking in (28) selects it as the optimal form; instead, they predict ill-formed (28a-ii) or (28b-i).

Instead of the reranking approach, Fukazawa (1998) presents a model that is crucially based on a concept of faithfulness developed in Correspondence Theory (McCarthy and Prince 1995; Benua 1995, 1997; Urbanczyk 1995, 1996). In Fukazawa's model, several stratum-specific faithfulness constraints are proposed to account for the overall organization of the stratified lexicon. Each morphological class is assumed to contain its own input-output correspondence relation: IO-Yamato, IO-Sino-Japanese, IO-Mimetic, IO-Foreign, and IO-Alien. That is, several sets of faithfulness constraints coexist in a single invariant ranking. The kind of invariant ranking assumed in the model can be synoptically illustrated as below. (M and F denote markedness and faithfulness constraints, respectively.)

(29)

$$\mathrm{M1} \gg \mathrm{F}x \gg \mathrm{M2} \gg \mathrm{F}y \gg \mathrm{M3} \gg \mathrm{F}z \begin{cases} \text{a.} \quad \mathrm{M1} \gg \boxed{\mathrm{F}x} \gg \mathrm{M2} \gg \mathrm{M3} & \text{(Lexical class } x\text{)} \\ \text{b.} \quad \mathrm{M1} \gg \mathrm{M2} \gg \boxed{\mathrm{F}y} \gg \mathrm{M3} & \text{(Lexical class } y\text{)} \\ \text{c.} \quad \mathrm{M1} \gg \mathrm{M2} \gg \mathrm{M3} \gg \boxed{\mathrm{F}z} & \text{(Lexical class } z\text{)} \end{cases}$$

Note that each of the rankings in (29) is implied by the single overall ranking: M1 » F*x* » M2 » F*y* » M3 » F*z*. Given this unified ranking, the phonological diversity of the lexical classes *x*, *y*, and *z* is properly captured by means of the dominance relation between markedness (M) and faithfulness (F) constraints. For instance, while lexical class *x* is subject only to the highest ranked markedness constraint M1, lexical class *y* must obey M2 in addition to M1.

The model with stratum-specific IO-faithfulness constraints provides an adequate account for the empirical problem posed by hybrid compounds. Fukazawa (1998) demonstrates that the constraint ranking in (30) explains the phonological structure of hybrid compounds.

(30)

$$\left(\begin{array}{l} \text{Ident (voi)-IO-SJ} \\ \text{Ident (voi)-IO-F} \\ \text{Ident (voi)-IO-A} \end{array} \right) \gg \quad \text{*NT} \quad \gg \left(\begin{array}{l} \text{Ident (voi)-IO-Y} \\ \\ \text{Ident (voi)-IO-M} \end{array} \right)$$

The faithfulness constraints Ident (voi)-IO are relativized to each stratum. Since postnasal voicing is respected in the Yamato and Mimetic strata, the Ident constraints relevant to these strata are dominated by the markedness constraint *NT. In contrast, postnasal obstruents in strata other than Yamato and Mimetic are exempt from voicing. This outcome is ensured by a constraint ranking in which the Ident constraints for Sino-Japanese, Foreign, and Alien are ranked above *NT. Given the ranking in (30), the well-formedness of the postnasal voicing patterns in the hybrid compound *tonbo-kenkyuuka* 'dragonfly-researcher' is correctly evaluated.[17] ("→" denotes the optimal output.)

(31) *tonbo-kenkyuuka* 'dragonfly-researcher'

/tonpo-kenkyuuka/		Ident (voi)-IO-SJ	*NT	Ident (voi)-IO-Y
a.	to**np**o-ke**nk**yuuka		**!	
b.	to**nb**o-ke**ng**yuuka	*!		*
→ c.	to**nb**o-ke**nk**yuuka		*	*
d.	to**np**o-ke**ng**yuuka	*!	*	

17 This analysis is based on Fukazawa, Kitahara, and Ota (1998: 51). Their PNV constraint is labeled *NT in order to maintain consistency of terminology in the present chapter.

The candidates (31b) and (31d), in which the Sino-Japanese element **kengyuuka* involves excessive voicing, are eliminated due to a violation of dominant Ident(voi)-IO-SJ. As for the remaining two candidates, (31c) is more preferable because it satisfies *NT better than (31a), which violates the constraint twice.

Since the leading idea was proposed by Fukazawa (1998), the optimality-theoretic model making use of stratum-specific faithfulness rankings has developed into the mainstream account for lexical stratification in Japanese. A number of studies have examined various kinds of phonological phenomena on the basis of the notion of multiple faithfulness relations. See Fukazawa, Kitahara, and Ota (1998, 1999, 2002), Fukazawa and Kitahara (2005), and Ito and Mester (1999, 2003, 2008) for further analyses and extended discussion.

5 The status of mimetics in the synchronic phonological lexicon

Let us return to mimetic phonology again. In this section, we will discuss the status of mimetics in the synchronic phonological lexicon of Japanese, referring to some descriptive findings and theoretical ideas presented in previous studies. The relationship between mimetics and Yamato will be a core topic in the discussion.

5.1 Affinity between mimetics and Yamato

As already discussed in section 3, the dual character of mimetics is clear when we compare them to Yamato with respect to their phonological properties. Discrepancies are observed between the Yamato and mimetic strata, even though both are included in a single etymological class indigenous to Japanese. For instance, Yamato items and mimetics behave identically with respect to Lyman's Law and postnasal voicing, but they exhibit totally opposite behaviors with respect to other kinds of phonological regularities, as summarized below.

Table 7: Phonological discrepancies between Yamato and mimetic strata

	Yamato	Mimetic
Lyman's Law	✓	✓
Postnasal voicing	✓	✓
Root-initial voicing	*	✓
Rendaku	✓	*
Singleton [p]	*	✓

(* = does not apply/prohibited, ✓ = applies)

Because of these discrepancies, the treatment of mimetics has been inconsistent in the previous Japanese linguistics literature. That is, it has been controversial whether mimetics should be regarded as an independent stratum from Yamato or not. This controversy is clearly due to the conspicuous *dissimilarities* listed in lower half of Table 7.

However, Hamano (2000) points out a few significant facts, listed in (32), indicating that mimetics have more phonological *similarities* with Yamato. (Each phonological property is labeled by the constraint that follows it in parentheses. For the sake of simplicity, we will refer each of the properties in (32) by means of this abbreviated, constraint-like notation in the following discussion.)

(32)	a.	Under-representation of the vowel /e/	(*[e])
	b.	Rhotic exclusion in root-initial position	(*#R)
	c.	Rarity of morpheme-medial [h]	(*. . .h. . .)
	d.	Minimal stem requirement based on a bimoraic template	(Stem=F)

All of these traits are shared by Yamato and mimetic items. First, /e/ is the least frequent vowel both in Yamato and in mimetic items. According to NLRI (1984: 25), vowel frequencies in bimoraic native morphemes are as follows: /a/: 831 > /i/: 636 > /o/: 559 > /u/: 500 > /e/: 446. Data on vowel frequencies in mimetics, presented by Hamano (1998: 47), also show that /e/ is the least frequent vowel. In bimoraic mimetic roots that appear reduplicated, the numbers are as follows: /a/: 198 > /o/: 174 > /u/: 161 > /i/: 140 > /e/: 59. Second, [r] is a quite peculiar segment in that it hardly ever appears in root-initial position. There are almost no words beginning with [r] in the Yamato vocabulary (NLRI 1984: 25), and the same goes for mimetics (Hamano 1998: 41). Recall that the data already given in the Tables 4 and 5 show this point clearly. (Although *p* is highlighted in the tables, it is the figures for *r* that are of interest here.) Third, morphemes containing [h] in intervocalic (i.e., morpheme-medial) position are extremely rare both in Yamato items and in mimetics. In Yamato items, intervocalic [h] is found only in a few exceptional words such as *ahureru* 'to overflow', *ahiru* 'duck', *yahari* 'likewise', and in a small number of etymologically reduplicated nouns such as *haha* 'mother' and *hoho* 'cheek'.[18] Intervocalic [h] in mimetics is found only in expressions that imitate coughing, hawking or laughing, all of which are related to "laryngeal or guttural sounds" as discussed in Hamano (1998: 145). Mimetic words such as *goho-goho* 'coughing', *ehen*/*ohon* 'sound of clearing one's throat', or *ahaha* 'laughing' are representative examples. And fourth, the bimoraic minimal stem template plays a pivotal role in prosodic morphology both in Yamato items and in mimetics. Poser (1990) examines a number of

18 The under-representation of intervocalic [h] in Yamato vocabulary is due to historical changes that affected [p]. See Takayama (this volume) for details.

foot-based phenomena to demonstrate the crucial role of the bimoraic foot in Japanese. Hypocoristic formation is a typical phenomenon of this kind. For example, the female name *Mariko* can be truncated either as *Mari*(*-tyan*), *Maa*(*-tyan*), *Mako*(*-tyan*), or *Riko*(*-tyan*), all of which consistently are two-moras long.[19] (The suffix *-tyan* is a kind of diminutive frequently attached to hypocoristic forms.) A large part of mimetic morphology is also subject to the same restriction, i.e., the minimum stem requirement, as discussed in Hamano (1998: 29–32). For example, monosyllabic mimetic roots such as *po-* or *gu-* cannot be used as proper stems; they always appear with a coda consonant or a glide (e.g., *pon* 'tapping', *gui* 'with a strong pull') to fulfill the bimoraic foot template.

On the basis of these facts, it is safe to say that Yamato items and mimetics are not completely different in nature. Indeed, they are basically homogeneous with respect to the majority of phonological properties, even though a few dissimilarities are observed. As Hamano (2000: 219) mentions, "the Yamato and the sound-symbolic strata have a long history of sharing many phonological properties."

To account for this kind of somewhat complicated but interesting relationship between Yamato and mimetics in the synchronic phonological lexicon, the core-periphery model discussed in section 4.1 provides an appropriate perspective. With reference to the "constraint map" presented by Ito and Mester (1995a: 834), the distribution of the constraint domains concerning each of the phonological regularities discussed above can be illustrated as in Figure 4. Approximate definitions of some of the constraints (*#R, *[e], *…[h]…, Stem=F) are given in (32) above.

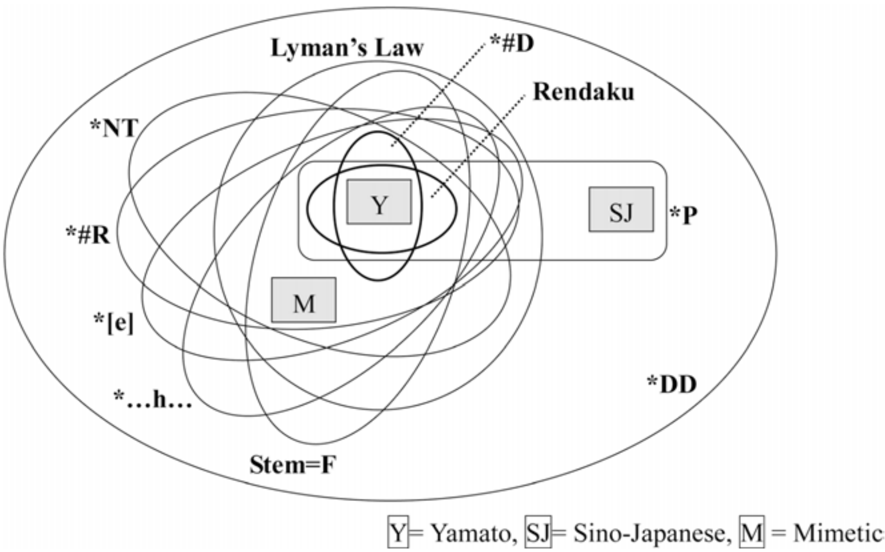


Figure 4: Constraint domains covering the Yamato and mimetic lexical areas

19 As demonstrated in Poser (1990) and Ito (1990), bimoraic truncation is applied not only to Yamato words but also to loanwords.

The whole lexical area containing Yamato, Mimetics, and Sino-Japanese is enclosed by the outermost constraint domain *DD (No voiced geminate). Rendaku and *#D (No root-initial voiced obstruent) occupy the core area of the phonological lexicon, and they cover only Yamato items. The domain defined by *P (No singleton [p]) encloses Yamato and Sino-Japanese together.

The most significant point to note here is that this constraint map shows that the domains covering mimetics are closely tied to the Yamato lexical area. All six constraint domains enclosing mimetics (Lyman's Law, *NT, *#R, *[e], *…[h]…, Stem=F) also enclose Yamato items but exclude Sino-Japanese items.[20] The simplified diagram below represents the affinity between Yamato and mimetics more clearly.

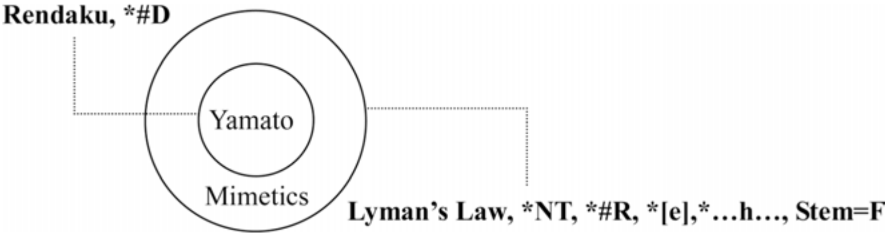


Figure 5: Relationship between Yamato and mimetics in the phonological lexicon

The concentric circles in Figure 5 correspond to the lexical areas of items indigenous to Japanese. Yamato (the non-mimetic native class) is much closer to the core, whereas the mimetic group occupies the outer area delimited by several phonological constraints other than those defining the inner domain. This configuration demonstrates that lexical items categorized as mimetic share some but not all phonological regularities with items classified as Yamato.

5.2 Double voicing and singleton [p]

One thing that remains elusive in the constraint map in Figure 4 is the domain delimited by the *P constraint. While Yamato is grouped together with Sino-Japanese in terms of *P, mimetics are excluded from this domain. Thus, one could say that Yamato and mimetic items are clearly distinguished from each other with respect to the *P constraint. However, we should not overlook a notable phenomenon which implies that the *P constraint is active even in mimetics. Although singleton [p] is a legitimate and rather frequent segment in mimetics, as discussed in section 3.4,

20 According to the constraint map presented by Ito and Mester (1995a: 834), Lyman's Law is regarded as a constraint that holds only in Yamato. However, as seen in (11), it exerts a profound effect on voicing patterns in mimetic items as well.

mimetic forms such as those in (33a) are not grammatical.[21] Corresponding well-formed items are shown in (33b).

(33)	a.	**depu-*	b.	*debu-*	'fatty, plump'
		**dapa-*		*daba-*	'loose, watery'
		**dapo-*		*dabo-*	'loose, big'
		**dapu-*		*dabu-*	'loose, baggy'
		**dopo-*		*dobo-*	'splashing'
		**dopu-*		*dobu-*	'mud splash'
		**zapa-*		*zaba-*	'washing, splashing'
		**zapu-*		*zabu-*	'washing, splashing'
		**zupa-*		*zuba-*	'boldly, frankly'
		**zupo-*		*zubo-*	'piercing, sinking'
		**zupu-*		*zubu-*	'sink into'
		**gepo-*		*gebo-*	'belching'
		**gepu-*		*gebu-*	'belching'
		**gapa-*		*gaba-*	'too large'
		**gapo-*		*gabo-*	'slurping'
		**gapu-*		*gabu-*	'gulping'
		**gopo-*		*gobo-*	'bubbling'

The thing to be noticed is that all the existing roots in (33b) exhibit a consistent behavior in terms of voicing. Not only the root-initial consonant, but also the second consonant of each root is voiced, and the second voiced obstruent is the bilabial stop [b] in each case (Hamano 1986, 1998, 2000; Nasu 1999). That is to say, the well-formed roots (33b) all violate the restriction against double obstruent voicing, namely, Lyman's Law, which is otherwise a rigid restriction in Yamato items and mimetics. Thus, the forms in (33) are quite paradoxical with respect to this restriction. The doubly-voiced items, which violate Lyman's Law, are well-formed, whereas the items that obey Lyman's Law are ill-formed.

The point to notice here is that singleton [p] in (33a) is the primary factor in this ill-formedness. The set of forms in (33) indicates that it is not the case that singleton [p] appears with no restrictions in the mimetic vocabulary. Although the double voicing in (33b) appears surprising at first sight, it has the consequence of avoiding singleton [p]. In other words, singleton [p] is a latently marked segment in mimetics,

---

21 In fact, one comes across a number of colloquial mimetic expressions containing forms such as those in (33a) in a quick Internet search. For example, there were 38,300 hits for *gapo-gapo* (ガポガポ) 'oodles and oodles' in the results of one Google search (May 4, 2013, 4:19 P.M. JST). However, such forms cannot be recognized as orthodox patterns appearing in conventional mimetic expressions. Dictionary entries can be one of source of evidence for the ill-formedness of the forms in (33a). In *Nihon kokugo daijiten* (Shogakukan 1979–1981), reduplicated mimetic words like those in (33a) are not listed, with the sole exception of *gopo-gopo* 'lightly bubbling'.

and it is prevented from emerging at the cost of violating Lyman's Law. In this respect, mimetics and Yamato items partially share a phonological regularity that treats singleton [p] as a marked segment.

However, it is appropriate to ask why these two lexical classes do not pattern together with regard to the strategy that prevents singleton [p] from emerging. While voicing takes place in mimetics, it does not serve as a strategy for prohibiting singleton [p] in Yamato items. In lieu of voicing, underlying /p/ in Yamato items is subject to debuccalization (*p~h* alternation), resulting in the glottal fricative [h] in output forms, as exemplified in (17a). In contrast, mimetic [p] never undergoes debuccalization; a form such form as **zuha-zuha* cannot be derived from the corresponding voiceless base *supa-supa* 'chopping-up'.

5.3 Phonology in morpheme-initial position

The discrepancy between Yamato and mimetic items emerges more explicitly in morpheme-initial position. Recall that [h] and [p] show quite opposite behaviors in their frequency of appearance in morpheme-initial position in Yamato items, as already seen in Table 5. While [h] is one of the most frequent segments in initial position in Yamato morphemes, [p] hardly ever appears in that position. In contrast, [p] contrasts with [h] in morpheme-initial position in mimetic items, as seen in minimal pairs such as *poro-poro* 'crumbly' vs. *horo-horo* 'shedding teardrops'. Although generally prohibited in the intervocalic position of voiced roots, as shown in (33), [p] appears quite freely in morpheme-initial position in mimetic roots. Interestingly, if Yamato morpheme-*initial* position and mimetic morpheme-*medial* position are compared with respect to the segments of interest, a quite systematic distributional regularity emerges. There is a complementary distribution with respect to the (il)licit segments in Yamato morpheme-initial position mimetic morpheme-medial position.

Table 8: Complementary distribution with respect to [h], [p], and [b]

Yamato		Mimetics	
(morpheme-*initial*)		(morpheme-*medial*)	
[h]	(*himo* 'string')	*[h]	(**suha*-, **zuha*-)[22]
*[p]	(**pimo*)	[p]	(*supa*- 'chopping-up')
*[b]	(**bimo*)	[b]	(*zuba*- 'boldly, frankly')

22 For the sake of simplicity, the few exceptional roots imitating laryngeal/guttural sounds (mentioned in section 5.1) are excluded from consideration here.

What is of great interest is that a similar kind of regularity holds in the distribution of voiced obstruent as well. As discussed in section 3.1, Yamato morphemes beginning with a voiced obstruent are relatively unusual, but voiced obstruents are generally favored in medial position. The data published by NLRI (1984: 25) demonstrate this asymmetry clearly. The ratio of voicing in initial obstruents in Yamato morphemes is only 15.1%, as seen in the following table. (The table rearranges the NLRI data for the sake of the present discussion.)

Table 9: Distribution of obstruents in bimoraic Yamato morphemes

Morpheme-*initial* position				Morpheme-*medial* position			
Voiced		Voiceless		Voiced		Voiceless	
b	36	h	132	b	80	h	15
d	25	t	125	d	50	t	122
z	15	s	128	z	60	s	96
g	20	k	153	g	74	k	124
96 (15.1%)		538 (84.9%)		264 (42.5%)		357 (57.5%)	
634 (100.0%)				621 (100.0%)			

In contrast, root-initial voicing is overwhelmingly favored in mimetics, as discussed in section 3.2. Recall that the data in Table 2 demonstrate this point. The ratio of voicing in root-initial obstruents is 45.0% in mimetics, and any kind of obstruent can be voiced in root-initial position. Thus, the distributional pattern of voiced obstruents in mimetics is the mirror image of that in Yamato items.

On the basis of the discussion thus far, a reasonable generalization emerges if we confine our attention to the regularity observed in morpheme-initial position. While both [p] and underlying voiced obstruents are licit segments in morpheme-initial position in mimetic items, neither of them is permitted in the case of Yamato items. This contrast is summarized below.[23] ("#" and "D" denote morpheme-initial position and a voiced obstruent, respectively.)

Table 10: Phonological regularity in the morpheme-initial position

	Mimetics	Yamato
#[p]...	✓	*
#D...	✓	*

(✓ = licit / * = illicit)

23 Needless to say, singleton [p] is uniformly illegitimate in Yamato items, regardless of its position. But recall that the *p~h* alternation occurs overwhelmingly in initial position, as discussed in section 3.3.

This generalization demonstrates that singleton [p] and voiced obstruents serve jointly as features that draw a line between Yamato and mimetic items. Although they are different types of segments, they pattern together with respect to their properties in morpheme-initial position.

5.4 Two sides of the same coin

The combined regularity shown in Table 10 is a key to understanding the relationship between Yamato and mimetic items in the synchronic lexicon of Japanese. Morpheme-initial [p] and initial obstruent voicing are closely related to each other in terms of their sound-symbolic functions in the language, in particular in mimetic items. As observed in the minimal pairs in (34), the phonological contrast between voiceless [p] and voiced [b] in root-initial position yields semantic contrasts between the reduplicative mimetics in (34a) and (34b).

(34)	a.	*pata-pata*	'pattering'	b.	*bata-bata*	'floundering'
		piri-piri	'smarting'		*biri-biri*	'numbed'
		pura-pura	'swinging'		*bura-bura*	'dangling'
		pera-pera	'rattling off'		*bera-bera*	'talking glibly'
		poro-poro	'crumbly'		*boro-boro*	'crumbling to pieces'

Recall that the same kind of phono-semantic contrast operates in all other types of obstruents as well. As already seen in (10), initial obstruent voicing plays a pivotal role in yielding phono-semantic contrasts in the mimetic vocabulary; morpheme-initial voicing symbolizes such meanings as heaviness, largeness, coarseness, or thickness with respect to the states of objects or manners of movement, as discussed in Hamano (1986, 1998). Because of this phono-semantic role, mimetics are exempt from the restriction against root-initial voicing. Instead, since mimetics are fundamentally sound-symbolic, root-initial voicing takes priority over the restriction.

In contrast, the restriction against root-initial voicing in the Yamato vocabulary has a close relationship with a familiar voicing regularity, namely, rendaku. Komatsu (1981: 104–107) makes an important observation that a voiced obstruent due to rendaku serves as a marker to indicate the morpheme boundary in compound words. He argues that it is precisely because the restriction against initial voicing is at work that the rendaku can function as a boundary marker in the Yamato stratum. This explanation is quite suggestive in that rendaku and the restriction against root-initial voicing can be grouped together as an integrated regularity. On the basis of this explanation, the exact opposite behaviors of Yamato and mimetic items, shown in the lower half of Table 7, can be easily accounted for. On the one hand, rendaku applies to Yamato items since the restriction against initial voicing is at work. On the other hand, rendaku does not take place in mimetic items, which are exempt

from the restriction against initial voicing. In the latter, if rendaku applied and yielded forms like **kata-gata*, it would ruin the phono-semantic voicing contrast in morpheme-initial position.

Based on the discussion thus far, the relationship between Yamato and mimetic items can be compared to the two sides of a coin. Heads (the Yamato stratum) exhibits what tails (the mimetic stratum) does not, and vice versa, even though they are parts of the same entity in which a number of basic phonological properties are shared. Indeed, while exhibiting a few opposing traits, Yamato and mimetic items share many more similarities, which make up the bulk of the "coin". This situation is summarized in the following table.

Table 11: Phonological (dis)similarity between Yamato and Mimetics[24]

	Yamato	Mimetics
Rendaku	✓	*
*#D	✓	*
*#P	✓	*
Lyman's Law	✓	✓
*NT	✓	✓
*#R	✓	✓
*[e]	✓	✓
*...h...	✓	✓
Stem=F	✓	✓

(✓ = obeyed, * = violated)

The point to notice is that the dissimilarity between Yamato and mimetic items emerges only in the properties related to phonological contrasts in morpheme-initial position (*#D and *#P). Continuing with the coin metaphor, the phonological differences listed in the upper half of Table 11 can be illustrated as below. (Rendaku is not included in the figure since it can be unified with *#D, following Komatsu's (1981) observation.)

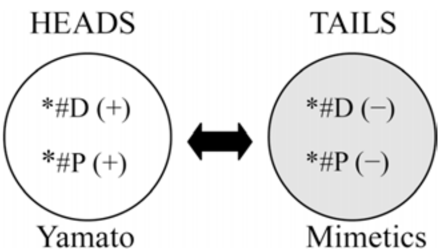


Figure 6: Opposite phonological traits that emerge in morpheme-initial position

24 The phonological properties are indicated by means of the constraints utilized in the Figure 4, except for "*#P," which denotes a ban on singleton [p] in morpheme-initial position.

Heads (Yamato items) and tails (mimetics) exhibit opposite values with respect to *#D and *#P, but these features and their opposite values are governed by a consistent principle: *preservation of phonological contrast in morpheme-initial position*. That is, whether or not a phonological contrast has to be maintained in morpheme-initial position is the key to deciding which side of the "coin" emerges.

5.5 Further issues

It must be noted, however, that the dual character described above does not directly correspond to a simple etymological categorization. Stated more explicitly, the border between Yamato and mimetics is not as clear-cut as might be expected but in fact quite fuzzy. Indeed, "tails" can appear even when the item of interest is originally a non-mimetic Yamato form. For example, the *#D constraint (the restriction against morpheme-initial voicing) is deliberately violated in non-mimetic words such as *bokeru* 'to become senile' (cf. *hokeru* 'be lost in thought'), *bareru* 'to be revealed, to transpire' (cf. *hareru* 'to clear up'), *goneru* 'to complain' (cf. *koneru* 'to argue for'), and *zama* 'messy appearance' (cf. *sama* 'appearance'), all of which convey a negative or pejorative nuance. These examples demonstrate that the morpheme-initial voicing is not necessarily an exclusive property of mimetics, but it is also utilized to create new word forms even in the Yamato vocabulary. As Komatsu (1981: 87–100) argues, morpheme-initial voicing generally serves to signal the negative character of the referent, and the Yamato and mimetic strata share this strategy.

Thus, it is not adequate to regard Yamato and mimetics as completely separate classes. Rather, they should be regarded as a unified category forming a harmonious whole in the phonological lexicon of Japanese. That is, the phonological properties (or lexical status) of the two categories are not independent of each other but instead, continuous with each other.

The same kind of dual behavior is observed with respect to the lexical usage of native items as well. There are in fact some native roots which can serve either as Yamato items or as mimetic items in the synchronic lexicon. For instance, according to Yamaguchi (2003: 188), the native root *kog(a)-*, which serves as a part of the stem in the Yamato verb *kogas-u* 'to burn, to fry', also serves as a constituent of the mimetic word *kongari* 'perfectly fried'.[25]

To capture the continuous character of native items in the phonological lexicon, the core-periphery model equipped with the notion of "constraint domain" seems to

25 Martin (1952: 68) also makes this point, remarking that "others, such as ko(.n.)ga.ri '(burnt) brown', yu(.k.)ku.ri 'slowly', seem to have a verbal base: koga.s.u 'scorches, chars', yuk.u 'goes' (a literary and dialect variant of ik.u, which occurs in standard speech only in a few forms like -yuk.i 'bound for')."

be on the right track. As illustrated in Figure 5, Yamato and mimetic items constitute adjacent and continuous domains in which some phonological constraints overlap. The domain in which an item of interest is included is not determined by its etymological origin but by the phonological properties of that item. Some items jump out of the inner domain into the outer domain, even though they are not etymologically mimetic morphemes. The Yamato words that show morpheme-initial voicing (e.g., *zama* 'messy appearance' < *sama* 'appearance') are representative examples of such behavior. Providing a more explicit account of this sort of continuous relationship (or dual behavior) between the Yamato and mimetic strata will be an important topic in future investigations of the synchronic configuration of phonological lexicon in Japanese.

Acknowledgments

Part of this work is supported by a Grant-in-Aid for scientific research (C) from Japan Society for the Promotion of Science (No. 23520455 and 26370437). I would like to express my gratitude to Timothy J. Vance, Donna Erickson, and Mitsuhiko Ota for their invaluable comments and suggestions.

References

Benua, Laura. 1995. Identity effects in morphological truncation. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18), 77–136. Amherst, MA: University of Massachusetts, GLSA.

Benua, Laura. 1997. *Transderivational identity: Phonological relations between words*. Amherst, MA: University of Massachusetts dissertation.

Fukazawa, Haruka. 1998. Multiple input-output faithfulness relations in Japanese. Ms., University of Maryland, College Park. [Available on Rutgers Optimality Archive, ROA-260-0598, http://roa.rutgers.edu]

Fukazawa, Haruka and Mafuyu Kitahara. 2005. Ranking paradoxes in consonant voicing in Japanese. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 105–121. Berlin and New York: Mouton de Gruyter.

Fukazawa, Haruka, Mafuyu Kitahara and Mitsuhiko Ota. 1998. Lexical stratification and ranking invariance in constraint-based grammars. In M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson and Tamra Wysocki (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 34-II*, 47–62.

Fukazawa, Haruka, Mafuyu Kitahara and Mitsuhiko Ota. 1999. Tajū seiyaku shōgō to goisō no kakutoku: Nihongo ni okeru shōko [Multiple faithfulness in the structure and acquisition of lexical strata: Evidence from Japanese]. *On'in Kenkyū* [Phonological Studies] 2. 77–84.

Fukazawa, Haruka, Mafuyu Kitahara and Mitsuhiko Ota. 2002. Constraint-based modeling of split phonological systems. *On'in Kenkyū* [Phonological Studies] 5. 115–120.

Hamada, Atsushi. 1960. Rendaku to renjō [Rendaku and renjo]. *Kokugo Kokubun* 29(10). 1–16.

Hamano, Shoko. 1986. *The sound-symbolic system of Japanese*. Gainesville, Florida: University of Florida dissertation.
Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford, Calif.: CSLI and Tokyo: Kurosio.
Hamano, Shoko. 2000. Voicing of obstruents in old Japanese: Evidence from the sound-symbolic stratum. *Journal of East Asian Linguistics* 9(3). 207–225.
Hashimoto, Shinkichi. 1938. Kokugo on'in no hensen [Historical development of Japanese phonology]. *Kokugo to Kokubungaku* 15(10). 3–40.
Hashimoto, Waka. 2006. Asahi shinbun shasetsu no gairaigo [Loanwords in the Asahi Shimbun editorials]. *Dōshisha Kokubungaku* 64. 178–186.
Hattori, Shiro. 1979. *Shinpan On'inron to seishohō* [Phonology and orthography, new edition]. Tokyo: Taishukan.
Irwin, Mark. 2011. *Loanwords in Japanese*. Amsterdam and Philadelphia: John Benjamins.
Ishiwata, Toshio. 1960. Zasshi no naka no gairaigo [Loanwords in magazines]. *Gengo Seikatsu* 161. 33–39.
Ito, Junko. 1990. Prosodic minimality in Japanese. In Michael Ziolkowski, Manuela Noske and Karen Deaton (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.
Ito, Junko and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17(1). 49–73.
Ito, Junko and Armin Mester. 1995a. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Oxford: Blackwell.
Ito, Junko and Armin Mester. 1995b. The core-periphery structure of the lexicon and constraints on reranking. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18), 181–209. Amherst, MA: University of Massachusetts, GLSA.
Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takeshi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin and New York: Mouton de Gruyter.
Ito, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Malden, MA. and Oxford, U.K.: Blackwell.
Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge, MA: MIT Press.
Ito, Junko and Armin Mester. 2008. Lexical classes in phonology. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 84–106. Oxford: Oxford University Press.
Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Ito, Junko, Armin Mester and Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26(4). 571–613.
Itō, Masamitsu. 2003. Kayō kyoku no naka no gairaigo gaikokugo [Loanwords and foreign words in pop songs]. *Nihongogaku* 22(7). 40–48.
Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology.* Santa Cruz, California: University of California dissertation.
Kindaichi, Haruhiko. 1976. Rendaku no kai [On explaining rendaku]. *Sophia Linguistica* 2. 1–22.
Komatsu, Hideo. 1981. *Nihongo no sekai 7: Nihongo no on'in* [The world of Japanese 7: Japanese phonology]. Tokyo: Chūōkōronsha.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kuroda, S.-Y. 1967. Sokuon oyobi hatsuon ni tsuite [On moraic obstruents and nasals]. *Gengo Kenkyu* 50. 85–99.

Kurisu, Kazutaka. 2000. Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9(2). 147–185.

Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Lyman, Benjamin Smith. 1894. The change from surd to sonant in Japanese compounds. *Oriental Studies*. 160–176. Philadelphia: Oriental Club of Philadelphia.

Martin, Samuel E. 1952. *Morphophonemics of standard colloquial Japanese* (Language Dissertation 47). Baltimore, Md.: Waverly Press.

McCarthy, John J. and Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18), 249–384. Amherst, MA: University of Massachusetts, GLSA.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Motoori, Norinaga. 1822. *Kojiki den* [Commentary on the *Kojiki*]. Edo: Maegawa Rokuzaemon.

Nakagawa, Yoshio. 1966. Rendaku, rensei (kashō) no keifu [The history of sequential voicing and (as a provisional term) sequential voicelessness]. *Kokugo Kokubun* 35(6). 302–314.

Nakata, Norio (ed.). 1982. *Nihongo no sekai 4: Nihon no kanji* [The world of Japanese 4: Chinese characters in Japan]. Tokyo: Chūōkōronsha.

Nasu, Akio. 1999. Onomatope ni okeru yūseika to [p] no yūhyōsei [Voicing in onomatopoeia and the markedness of [p]]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 3(3). 52–66.

Nishimura, Kohei. 2003. Lyman's Law in loanwords. Nagoya University MA thesis.

Nishimura, Kohei. 2006. Lyman's Law in loanwords. *On'in Kenkyu* 9. 83–90.

Nishio, Toraya. 2002. Goshu [Word types]. In Michiaki Saito (ed.), *Goi imi* [Words and meaning], 29–53. Tokyo: Asakura Shoten.

NLRI (ed.). 1964. *Hōkoku 25: Gendai zasshi kyūjisshu no yōgo yōji III: Bunseki* [Report 25: Vocabulary and Chinese characters in 90 magazines of today, vol. III: Analysis]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

NLRI (ed.). 1984. *Goi no kenkyū to kyōiku (1)* [Research and teaching in lexicology (1)]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

NLRI (ed.). 1987. *Hōkoku 89: Zasshi yōgo no hensen* [Report 89: Changes in the language of a magazine]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

NLRI (ed.). 1990. *Gairaigo no keisei to sono kyōiku* [Word formation and teaching of loanwords]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

Okumura, Mitsuo. 1955. Rendaku. In Kokugo Gakkai (ed.), *Kokugogaku jiten* [Glossary of Japanese linguistics], 961–962. Tokyo: Tōkyōdō Shuppan.

Ota, Mitsuhiko. 2004. The learnability of the stratified phonological lexicon. *Journal of Japanese Linguistics* 20. 19–40.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Poser, William J. 1990. Evidence for foot structure in Japanese. *Language* 66(1). 78–105.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Ms., Rutgers University and University of Colorado.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Rice, Keren D. 1993. A reexamination of the feature [sonorant]: The status of 'sonorant obstruents'. *Language* 69(2). 308–344.

Rice, Keren D. 1997. Japanese NC clusters and the redundancy of postnasal voicing. *Linguistics Inquiry* 28(3). 541–551.

Rice, Keren D. 2005. Sequential voicing, postnasal voicing, and Lyman's Law revisited. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 25–45. Berlin and New York: Mouton de Gruyter.

Sakurai, Shigeharu. 1972. Heian Insei jidai ni okeru wago no rendaku ni tsuite [On rendaku in Yamato words in Heian and Insei period]. *Kokugo Kokubun* 41(6). 1–19.

Sato, Hirokazu. 1989. Fukugōgo ni okeru akusento kisoku to rendaku kisoku [Accent rules and rendaku rules in compound words]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 233–265. Tokyo: Meiji Shoin.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shogakukan (ed.). 1979–1981. *Nihon kokugo daijiten (shukusatsuban)* [Great Dictionary of the Japanese Language (abridged edition)]. Tokyo: Shogakukan.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Ueno, Kagetomi. 1980. *Eigo goi no kenkyū* [A study on English vocabulary]. Tokyo: Kenkyusha.

Umegaki, Minoru. 1963. *Nihon gairaigo no kenkyū* [A study on loanwords in Japanese]. Tokyo: Kenkyusha.

Urbanczyk, Suzanne. 1995. Double reduplications in parallel. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18), 499–531. Amherst, MA: University of Massachusetts, GLSA.

Urbanczyk, Suzanne. 1996. *Patterns of reduplication in Lushootseed*. Amherst, MA: University of Massachusetts dissertation.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Yamaguchi, Nakami (ed.). 2003. *Kurashi no kotoba: Gion-gitaigo jiten* [Words in everyday Japanese: Dictionary of onomatopoeic and mimetic words]. Tokyo: Kodansha.

Junko Ito and Armin Mester

7 Sino-Japanese phonology

1 Introduction

This chapter takes up the Sino-Japanese substratum of the Japanese lexicon (*kango* 漢語, henceforth SJ), historically the result of intensive borrowings from Chinese at different periods. Large-scale and systematic borrowing began in the pre-Nara period (6th century AD) in connection with the introduction of Buddhism, followed by a second period in the 8th century, and a third one in the 14th century closely associated with Zen. For the written language, this means that it is not unusual for a given Chinese character (*kanji* 漢字) to have, besides a native-Japanese reading (*kun-yomi* 訓読み), several different but similar Sino-Japanese readings (*on-yomi* 音読み) traceable to different times of borrowing, different stages and dialects of Chinese dominant at the time, and different adaptation strategies. Thus 京 'capital' can be read as *kyoo*, *kee*, and *kin*. The role of SJ in Modern Japanese is comparable in English to that of the distinct but massive borrowings of Latinate/Romance vocabulary following the Norman Conquest and of Greek vocabulary during the Renaissance. In terms of its contexts of use, the SJ vocabulary ranges from items found in informal everyday conversation to items exclusive to formal discourse (see Shibatani 1990: 142–147 and work cited there), and its size is considerable (ca. 60% of a modern dictionary, and 20% of ordinary speech). Surprisingly, despite its very long history within Japanese, SJ has preserved many of the characteristics that distinguish it from the other two major constituents of the Japanese lexicon (see Nasu, this volume, and Kubozono, Ch. 8, this volume), the native Japanese vocabulary (*wago* 和語, or *yamatokotoba* 大和言葉) and Western loans (*gairaigo* 外来語). Rather, SJ continues to form a separate lexical stratum with unique morpheme-structural, prosodic, and segmental characteristics (Martin 1952; McCawley 1968; Ito and Mester 1996) many of which can be traced back to the monosyllabic shape of the Chinese source words, with subsequent adaptations throughout the history of the Japanese language.[1]

The goal of this chapter is to outline these special phonological properties and alternations within the synchronic grammar of Japanese, summarize previous work, and sketch new developments towards a better understanding of both the segmental and prosodic properties characterizing SJ phonology. Section 2 introduces the main prosodic characteristics of SJ items in terms of root and word size restrictions, and gives an overview of the special segmental make-up of SJ roots. Section 3 is devoted to the phonology of SJ compounding: Compounding at the root level gives rise to an

---

1 See Nasu (this volume) for discussion of the core-periphery model (Ito and Mester 1995) of the stratified lexicon of Japanese, and see also Takayama (this volume) and the History Volume in the same series for details about the history of the language.

interesting set of phonological alternations due to root-syllable alignment conditions, and higher-level compounding at the word level provides evidence for crisp prosodic word edges, and shows how interface mapping constraints play a key role in the analysis of SJ compounding.

2 Root structure and segment distribution

2.1 Root and word size

Two closely related prosodic characteristics are crucial to the understanding of the phonological properties and alternations found in SJ items. The first is a limit on the size of SJ roots stated in (1),[2] unsurprising given their monosyllabic Chinese sources. Individual SJ roots are either a single light syllable (e.g., /ka/ 科 'field', /ki/ 気 'spirit'), a single heavy syllable (e.g., /kan/ 館 'building', /nai/ 内 'internal', /kuu/ 空 'void'), or two light syllables with a monosyllabic allomorph (e.g., /gaku/~/gak/ 学 'study, scholarship', /betu/~/bet/ 別 'separate'), corresponding to the (maximally) bimoraic foot shown to play a central role in Japanese morphology and phonology (see Ito and Mester, Ch. 9, this volume, and Otake, this volume).

(1) Root size: $|\ \text{root}_{\text{SJ}}\ | = \text{ft}$
$\text{ft} \leq 2\mu$, i.e., $\text{ft} = \sigma_\mu\sigma_\mu,\ \sigma_{\mu\mu},\ \sigma_\mu$

We return later to the question of why this prosodic limit needs to be stated in terms of the phonological foot and not directly in terms of moras.

The second characteristic, less discussed but nevertheless also critical for the proper understanding of both accentual and segmental alternations, is the size of words composed of SJ roots. Most SJ roots occur only in combination with other SJ roots (e.g., *dai+gaku* 大学 'great+study, university', *gaku+nai* 学内 'study+inside, school-internal', *gaku+see* 学生 'study+person, student', *sen+see* 先生 'previous+person, teacher'), not in isolation or compounded with items from the rest of the lexicon. A close parallel in English is the combinatorics of Greek roots (e.g., *cosmo+logy, micro+cosm, helico+pter, ptero+saur*, etc.), which are mostly not independent words and combine overwhelmingly only with other Greek roots.[3] Since SJ roots rarely

2 The notation |x| refers to the prosodic size of element x.

3 Hybrids do exist, such as /ba+syo/ 場所 'place' (native+SJ) or /zyuu+bako/ 重箱 'layered serving box' (SJ+native), (the latter is even used to refer to this kind of mixed formation in *juubako yomi* 重箱読み 'mixed reading'). However, it is our impression that their number is smaller than that of corresponding hybrids of Latin and Greek morphemes in English, which are rather frequent, perhaps because both are loans whose exact etymological pedigree is not clear to many users of English: E.g., *sociology* from the Latin *socius* 'comrade' and the Greek λόγος (logos) 'reason', or *television* from the Greek τῆλε (tēle) 'far' and the Latin *visio* 'seeing'.

occur in isolation, the roots themselves are only listed in specialized SJ root dictionaries (*kanwa-jiten* 漢和辞典). Regular Japanese dictionaries (*kokugo-jiten* 国語辞典) list independently occurring morphologically complex SJ lexical items composed of two SJ roots, which, given the root size restriction, are prosodically two feet.

(2) Word size: $|\, \text{word}_{SJ} \,| = |\, \text{root}_{SJ} + \text{root}_{SJ} \,| = 2\text{ft}$

As we will see below in section 3, these prosodic characteristics, both root size and word size, turn out to have implications for the realization of SJ items, with theoretical consequences for our understanding of the way prosodic structure is regulated.

2.2 Segmental composition

In addition to their prosodic size limit, SJ roots are highly restricted in their segmental composition, as shown in (3)–(5).[4]

(3) a. CV

ka	科	'department'
i	胃	'stomach'
gu	具	'material'
ke	家	'house'
ko	古	'old'

b. CVV

bee	米	'rice'
kyoo	京	'capital'
huu	風	'wind'
dai	大	'big'
sui	水	'water'

c. CVN

kon	今	'this'
ken	県	'prefecture'
kan	完	'complete'
kin	金	'money'
gun	軍	'army'

The basic generalization for CVV roots is that V_2 must be a high vowel, /i/ or /u/. The two sequences */ii/ and */oi/ are excluded, and most of the remainder are subject to

4 Since onsets are optional, "CV" in what follows should be understood as comprising both CV and V, etc.

monophthongization (4) (or gliding+compensatory lengthening, see Poser 1988 and Kubozono, Ch. 5, this volume).

(4)	iu	→	yuu	*ii		
	eu	→	yoo	ei	→	ee
	au	→	oo	ai	→	ai
	ou	→	oo	*oi		
	uu	→	uu	ui	→	ui

As a result, besides the two diphthongs /ai/ and /ui/, SJ has the three long vowels /ee/, /oo/, and /uu/, but no */aa/ or */ii/. The situation contrasts both with the native stratum, where long vowels are rare, and with the stratum of Western loans, where all five long vowels are common (see Moreton, Amano, and Kondo 1998 and Kubozono, Ch. 8, this volume). Several experiments (Moreton and Amano 1999; Gelbart and Kawahara 2007) have shown these distinctions between lexical strata (in particular, the absence of /aa/ in SJ vs. its presence elsewhere) to be psychologically real.

The only possible word-final consonant in Japanese is the moraic nasal (realized with dorso-uvular closure, see Vance 2008: 99–100), transcribed with *n* in (3c), and it is the only final consonant in monosyllabic SJ roots. Chinese, however, especially the historical varieties that SJ is based on, allowed a larger range of coda consonants including voiceless plosives, and source items with such codas gave rise to SJ items as in (5), which occur either as disyllabic CVCV or as monosyllabic CVC, depending on the phonological context (as explored below in section 3). We represent SJ roots in accordance with a conventional view of their underlying representations, in effect close to the *kunrei* style of transliteration (so /y/ indicates the palatal glide, etc.). Slashes (/…/) indicate underlying forms in the sense of generative grammar, not structuralist phonemic transcriptions. For example, /yaku/ corresponds to phonetic [jakɯ], /butu/ to [bɯtsɯ], /hati/ to [hatʃi], /tyaku/ to [tʃakɯ], etc., see sections 2 and 3 of the introduction by Kubozono (this volume).

(5) a. CVtu

atu	圧	'press'
betu	別	'different'
hitu	筆	'writing'
butu	物	'thing'
sotu	卒	'graduate'

b. CVti

hati	八	'eight'
kiti	吉	'luck'

c. CVku

iku	育	'be raised'
tyaku	着	'arrival'
huku	服	'luck'
hoku	北	'north'

d. CVki

teki	敵	'enemy'
riki	力	'power'

C_2 is always a voiceless stop (/t, k/), and V_2 is always a high vowel (/i, u/), as shown in (6).

(6) Segmental composition:

$$\begin{array}{cccc} C_1 & V_1 & C_2 & V_2 \\ & & | & | \\ & & \begin{Bmatrix} t \\ k \end{Bmatrix} & \begin{Bmatrix} u \\ i \end{Bmatrix} \end{array}$$

The C_2 restriction to voiceless plosives is inherited from the Chinese source words (the third voiceless stop /p/ was historically lost in Japanese), and the V_2 restriction is a reflection of the fact that this vowel was inserted in Japanese, due to a coda condition more stringent than the one found in the source dialect of Chinese. We can capture the V_2 restriction by means of a constraint on vowel sonority in weak positions of feet such as (7) (following de Lacy 2002: 118, see also de Lacy 2006).

(7) *NONHEAD(ft) ≥ {e, o}
Assign a violation for every foot nonhead that is equally or more sonorous than mid vowels (i.e., /e o a/).

It turns out, however, that the choice of V_2 is even more restricted: Not only is V_2 always high, its backness is also almost totally predictable from other properties of the form. The relevant generalizations are due to the study of Martin (1952), with further refinements in Tateishi (1990) and Ito and Mester (1996). The situation is summarized in (8).

(8) Cooccurrence table for V_2

V_1 = \ C_2 =	t	k
a	*u*†	*u*
o	*u*	*u*
u	*u*	*u*
e	*u*	*i*
i	*u*†	*i*/*u*††

†very few occurrences of *i* as V_2
††genuine contrast between *u* and *i* as V_2

In the overwhelming number of cases, V_2 is /u/: For *t*-roots, this is exceptionless when V_1 is /o/, /u/, and /e/, and there are only a handful of exceptions when V_1 is /a/ or /i/ (the number words /hati/ 八 'eight', /iti/ 一 'one', /siti/ 七 'seven', and two other isolated examples, /niti/ 日 'sun' and /kiti/ 吉 'good luck'). *k*-roots also show uniform /u/ after the back vowels /a, o, u/. After front vowels in V_1 position, there is something resembling a harmony pattern, as Tateishi (1990) has recognized: We find /i/ as the only option when V_1 = /e/ (e.g., /seki/ 石 'stone'), and as an option alongside /u/ when V_1 = /i/. The environment /*ik_* is therefore the only environment where a genuine contrast between /i/ and /u/ is found in V_2-position: Examples include /siki/ 式 'ceremony' vs. /ziku/ 軸 'axle', and /tiku/ 蓄 'accumulate' vs. /riki/ 力 'power'. The default color of the high V_2 vowel is thus [+back], i.e., /u/, arguably the unmarked vowel of the SJ and the Foreign lexical strata. Different from the native stratum, where /i/ is the prime candidate for the default vowel (see Poser 1984), the bulk of the /i/-cases in SJ arise through harmony, with [–back] harmony holding either uniformly or as a lexical option.

The almost total predictability of V_2 in SJ roots of the form CVCV implies that specifying this vowel in underlying representations is redundant and misses a major generalization. Earlier work therefore hypothesized that V_2 is underlyingly absent in all cases besides the exceptional cases involving /i/, and posits /bet/, /gak/, etc., as underlying representations. Under this view, vowel insertion is prosodic epenthesis triggered by an obstruent exclusively liked to the coda (Ito 1986; Tateishi 1990), a scenario which has become known in Optimality Theory (OT) as the coda condition (CodaCond). The default vowel /u/ is epenthesized to make forms like /bet/ syllabifiable, resulting in the disyllabic form /betu/.

In the current context, a number of possible analyses arise. Option 1 is to carry on with the traditional epenthesis analysis, positing monosyllabic SJ roots where V_2

is underlyingly absent and /bet/, /gak/, etc., are the unique underlying representations. For *k*-roots, but not *t*-roots, the epenthesized vowel agrees in backness with V_1 = e due to a constraint demanding harmony. Only truly exceptional cases (like /hat, hati/ 'eight') have both a monosyllabic CVC-allomorph and a disyllabic CVCV-allomorph. Option 2 is like Option 1, but with allomorph listing playing a larger role. V_2 = /u/ continues to be supplied by epenthesis to underlyingly monosyllabic roots. Allomorph listing applies to all roots with a monosyllabic CVC-form and a disyllabic CVCi-form, both to the unpredictable cases like /hat, hati/ 'eight', as well as the predictable backness harmony cases like /sek, seki/ 'stone'. Option 3 extends allomorph listing to all SJ roots and posits pairs of URs, /CVC, CVCi/ or /CVC, CVCu/, in all instances, treating the exceptional cases (like /hat, hati/ 'eight'), the backness-harmony cases (like /sek, seki/ 'stone'), and the regular cases with V_2 = /u/ (like /bet, betu/, /gak, gaku/), in the same way.

This does not exhaust all possible options, and here is not the place to argue for or against any of them, which differ mainly in terms of how much of the overall pattern they attempt to derive from general principles, potentially earmarked for the SJ vocabulary stratum. For concreteness, we will proceed by assuming Option 2, noting that this choice does nor rest on a principled argument (see Kurisu 2000 for more discussion of these issues, and a specific proposal).

3 Compounding and its prosody

3.1 Root compounding and alignment

As discussed in section 1, SJ roots occur mostly compounded with other SJ roots, and are listed in such collocations in regular dictionaries. The situation is similar to that of Greek roots in the English lexicon, where whole compounded forms (e.g., *pentagon*, *helicopter*) are listed in the dictionary as lemmas, not the individual roots (*penta-*, etc.). In both cases, the meanings are often non-compositional (e.g., /ben+kyoo/ 勉強 'effort-hard, to study', /sen+see/ 先生 'previous+born, teacher', or *helico+pter* 'curved+wing', *anthropo+logy* 'human-study', etc.). On the other hand, whereas in Greek compounds cross-morpheme syllabification *he.li.co+p.ter* and cross-morpheme footing (*anthro*)(*po+lo*)*gy* often make the prosodic boundaries of the two roots opaque, such cross-morpheme syllabification does not occur in SJ, even with vowel-initial morphemes as second members. Thus, /man/+/in/ 満員 'full capacity' and /gak/+/i/ 学位 'academic degree' do not appear as **ma.n+in* and **ga.k+i*, but as *man.+in* with a nasal coda, and as *ga.ku.+i* with its first member appearing in its vowel-final form allomorph to avoid an obstruent coda violation **gak.+i*. As a result, SJ root boundaries in compounds are impermeable to syllabification, and the two roots remain clearly recognizable as prosodic units in the output form.

The situation is undoubtedly at least in part influenced by the writing system, where each SJ root is represented by a single *kanji* and has a clear and separate orthographic identity. In phonological terms, it is unlikely that this behavior is due to cyclic syllabification, given that one of the most central results of Lexical Phonology is the very fact that roots do not constitute cyclic domains (see Brame 1974, Kiparsky 1982, and Inkelas 1989 for discussion and argumentation). The reason for the syllabic closure of SJ roots is rather to be sought elsewhere, viz., in the size restrictions governing SJ roots and words discussed in the previous section (and repeated here in (9)).

(9) a. Root size: $|\ \mathrm{root}_{SJ}\ | = \mathrm{ft}$
$\mathrm{ft} \leq 2\mu$, i.e., $\mathrm{ft} = \sigma_{\mu}\sigma_{\mu}, \sigma_{\mu\mu}, \sigma_{\mu}$

b. Word size: $|\ \mathrm{word}_{SJ}\ | = |\ \mathrm{root}_{SJ} + \mathrm{root}_{SJ}\ | = 2\mathrm{ft}$

Related to (9a), there are alignment constraints (interface mapping constraints) requiring root edges to match syllable edges (10), and it is these constraints that are responsible for the resistance to resyllabification.

(10) ALIGN-ROOT$_{SJ}$-TO-SYLLABLE
a. Align-Left (Root$_{SJ}$, Syllable)
b. Align-Right (Root$_{SJ}$, Syllable)

ALIGN-ROOT is a two-partite constraint governing both edges, left and right. It disallows syllabification across root boundaries by requiring root-initial elements to be syllable-initial (10a), henceforth ALIGN-ROOT-L(EFT), and root-final elements to be syllable-final (10b), henceforth ALIGN-ROOT-R(IGHT).[5] Both ALIGN-ROOT-L and ALIGN-ROOT-R are observed when the first root ends in a nasal (11a) or a vowel (11b). Alignment is satisfied when the syllable edge "." and the root edge "|" are not separated by segmental material.

(11)	a.	/san+po/	|.sam|.po.|	'scatter+walk', 'stroll'	散歩
		/san+koo/	|.saŋ|.koo.|	'go+think, reference'	参考
		/han+mee/	|.ham|.mee.|	'understand+light, reveal'	判明
		/han+bai/	|.ham|.bai.|	'trade+sell, sale'	販売
	b.	/koo+kan/	|.koo|.kan.|	'associate+replace, exchange'	交換
		/tai+kai/	|.tai|.kai.|	'big+event, convention'	大会

5 Constraints of this type were first explored in OT by Prince and Smolensky (1993 [2004]) and McCarthy and Prince (1993). The edge-based form of such interface constraints linking grammatical and prosodic categories was originated by Selkirk (1986).

The surface assimilation pattern (/np/ → [mp], /nk/ → [ŋk], etc.) still fulfills alignment in terms of the last segment of the root, and place-assimilated coda nasals are allowed in Japanese (Ito and Mester 1993; see also the introduction by Kubozono, this volume, and Ito and Mester, Ch. 9, this volume). Alignment is enforced even when the second root is vowel initial, and the resulting C.V juncture persists in the output form, as the examples in (12) show (there is no systematic insertion of a default phonetic onset filler, laryngeal ([ʔ]) or other; only the relevant medial junctural alignment is indicated here).[6]

(12)

	Aligned	Disaligned		
/sin+an/	sin|.an	*si.n|an	'new plan'	新案
/kan+i/	kan|.i	*ka.n|i	'simplicity'	簡易
/hon+ee/	hon|.ee	*ho.n|ee	'headquarters'	本営
/man+in/	man|.in	*ma.n|in	'full capacity'	満員

In OT terms, root alignment (10) outranks the onset requirement (ALIGN-ROOT$_{SJ}$ » ONSET), resulting in an internal open juncture utterly foreign to Yamato (native) items, where we find cross-morpheme syllabification to fulfill ONSET. For example, the final consonants in verb roots such as /kik/ 'to hear' and /tanom/ 'to request' syllabify with the vowel-initial suffixes (*ki.k+u*, *ta.no.m+u*), showing that no comparable root alignment is in force, undoubtedly also related to the fact that the native vocabulary knows no root-size restriction.

The merit in having separate alignment constraints for left and right root edges lies in the fact that the two are not equal in strength: Whereas ALIGN-ROOT-L is never violated, there are many violations of ALIGN-ROOT-R when the root is obstruent-final (i.e., *t*-final or *k*-final). When these obstruent-final roots are second members of compounds, epenthesis takes place to avoid a violation of CODACOND. Such epenthesis is disaligning, leading to an ALIGN-ROOT-R violation in (13), where the right edge of the root ("t|") is not at the right edge of the syllable ("*u*."). The alignment-violating forms surface as optimal because of the ranking CODACOND » ALIGN-ROOT. Combined with ALIGN-ROOT » ONSET established earlier, the overall ranking is CODACOND » ALIGN-ROOT » ONSET.

6 There is a historical linking pattern where the root-final consonant occupies both coda and onset position, surviving in contemporary Japanese in isolated forms such as: *ten.n|oo* 'emperor' from /ten+oo/ 天皇 'heavenly sovereign'; *un.n|un* from /un+un/ 云々 'various'; *gin.n|a*n from /gin+an/ 銀杏 'silver+fruit, gingko seed'. A similar linking pattern has been observed in recent loanwords (e.g., *pin.n|appu* for 'pin-up', *ran.n|awee* 'run-away', see Vance 2010), but there is no general tendency or variation in most SJ compounds (i.e., **sin.n|an*, **kan.n|i*, **hon.n|ee* are not possible output forms in (12)).

(13)

	ALIGN-ROOT-R violation	CODACOND violation		
/hai+tat/	haitat|*u*.	*haita*t*.|	'delivery'	配達
/dai+gak/	daigak|*u*.	*daiga*k*.|	'university'	大学

In (14), the initial SJ root is *t/k*-final. Syllabifying the root-final obstruent consonant as the onset of the second root violates ALIGN-ROOT-L (and in addition ALIGN-ROOT-R). However, since the root-final consonant cannot remain a coda because of CODACOND, the optimal output is one in which ALIGN-ROOT-R is violated.

(14)

	ALIGN-ROOT-R violation	ALIGN-ROOT-L violation	CODACOND violation		
/bet+en/	bet|*u*.en	*be.*t*|en	*be*t*.|en	'farewell dinner'	別宴
/gak+i/	gak|*u*.i	*ga.*k*|i	*ga*k*.|i	'academic degree'	学位
/kok+oo/	kok|*u*.oo	*ko.*k*|oo	*ko*k*.|oo	'king'	国王
/hat+an/	hat|*u*.an	*ha.*t*|an	*ha*t*.|an	'proposal'	発案

With *t*/*k*-final roots, epenthesis occurs not just with vowel-initial second members, but also with most consonant-initial second members because *t* and *k* cannot serve as codas, as shown in (15).

(15)

/bet+noo/	bet*u*.noo	*be*t*.noo	'separate payment'	別納
/bet+bin/	bet*u*.bin	*be*t*.bin	'separate carrier'	別便
/hat+den/	hat*u*.den	*ha*t*.den	'start electricity'	発電
/kat+yak/	kat*u*.yak*u*	*ka*t*.ya*k*	'live+leap, be active'	活躍
/gak+mee/	gak*u*.mee	*ga*k*.mee	'school name'	学名
/gak+gai/	gak*u*.gai	*ga*k*.gai	'outside of campus'	学外
/tok+bet/	tok*u*.bet*u*	*to*k*.be*t*	'special'	特別
/hak+too/	hak*u*.too	*ha*k*.too	'white sugar'	白糖
/gak+see/	gak*u*.see	*ga*k*.see	'student'	学生
/tok+tyoo/	tok*u*.tyoo	*to*k*.tyoo	'characteristics'	特徴
/hak+hat/	hak*u*.hat*u*	*ha*k*.ha*t*	'white hair'	白髪

The examples of *t*/*k*-final roots seen so far have all undergone epenthesis, raising the question of whether ALIGN-ROOT-R is ever active in the language at all. Obeying ALIGN-ROOT-R are the cases with nasal-final or vowel-final roots (see (11) above), where the constraint can be fulfilled without violating CODACOND. There is, however, one syllabic configuration where obstruent codas are allowed, namely, as the first part of a geminate, where coda licensing is not an issue (see Ito 1986, Goldsmith

1990, Ito and Mester 1993, and work cited there for justification and details). If a root-final /t/ and /k/ is followed by a root-initial /t/ and /k/, respectively, i.e., in the configurations /Vk+kV/ or /Vt+tV/, the obstruent codas are allowed because they are the first part of geminate structures (see Kawahara, Ch. 1, this volume, and Kawagoe, this volume). Right-alignment is fulfilled here, and resorting to epenthesis is not necessary. Unless other constraints (e.g., CODACOND) are at play, root alignment decides on the non-epenthetic candidate.[7]

(16)	…Vk+kV…	/gak+koo/	gak.koo	*gaku.koo	'school'	学校
		/hak+kot/	hak.kotu	*haku.koo	'skeleton'	白骨
		/gak+ki/	gak.ki	*gaku.ki	'musical instrument'	楽器
	…Vt+tV…	/bet+tak/	bet.taku	*betu.taku	'detached villa'	別宅
		/hat+tat/	hat.tatu	*hatu.tatu	'development'	発達
		/kot+too/	kot.too	*kotu.too	'antique'	骨董

Besides the identity cases, where underlying /CVk/ and /CVt/ surface as such, we also find underlying /CVt/ surfacing as /.CVs./, /.CVp./, and /.CVk./ before /s/, /p/, and /k/, respectively (17).[8]

(17)	t+p → pp	/bet+puu/	bep.puu	'letter under separate cover'	別封
		/bet+poo/	bep.poo	'different message'	別報
		/hat+pyoo/	hap.pyoo	'announcement'	発表
	t+s → ss	/hat+soo/	has.soo	'shipment'	発送
		/bet+sat/	bes.satu	'separate volume'	別冊
		/bet+syu/	bes.syu	'another kind'	別種
	t+k → kk	/bet+koo/	bek.koo	'separate clause'	別項
		/hat+ken/	hak.ken	'discovery'	発見
		/bet+kyo/	bek.kyo	'separation, limited divorce'	別居

7 This is where the precise analysis of the vowel-zero alternations in SJ roots becomes important. While alignment is violated when the underlying form is /CVC/, as we are assuming here, this is not so in an allomorphy analysis where both /CVC/ and /CVCu/ are available as underlying forms. In the latter case, while right alignment is not an issue, some other element in the analysis (such as an allomorph preference relation, or a syllable minimization constraint) has to force roots to ever appear in the CVC form.

8 /p/ at the beginning of the second root in such forms alternates with /h/ in other contexts: /bep+*p*oo/ 別報 'different+information, another report' vs. /*h*oo+koku/ 報告 'information + announce, report', etc., see Nasu (this volume) on the /h/~/p/ alternation.

Even though place and stricture features have changed, both alignment requirements still hold in these examples: Root-final segments are syllable-final, root-initial segments syllable-initial. The segment in the onset preserves its features because of positional faithfulness (IDENT-ONSET or IDENT-σ_1, see Beckman 1997, Casali 1997, and Lombardi 1999), but the coda *t* acquires the place and stricture features of the following consonant. Once it is the first part of a geminate, there is no need for epenthesis, and right-alignment is satisfied.

Feature-changing gemination of this type is only allowed between *t*-final roots and voiceless obstruent-initial roots, as listed in (17) above. The details of why feature-changing gemination does not apply to other combinations are explained by various other constraints governing the Japanese phonological system. Voiced geminates (voiced obstruent geminates like *gg*, *bb*, *dd*, and nonnasal sonorant geminates like *rr*, *ww*, etc.) as conceivable outcomes in examples like (15) violate a general constraint on Japanese syllable structure holding throughout the non-Foreign (i.e., Yamato, SJ, and Mimetic) vocabulary (see Kawahara, Ch. 1, this volume, Kawagoe, this volume, and Nasu, this volume). Nasal geminates (*nn*, *mm*) arise from assimilation with nasal-final roots (see (11) above), but not with *t*-final roots (/bet+noo/ appears as *betu+noo*, not **ben+noo*), hence feature-changing gemination must be restricted to obstruent-obstruent combinations, which indicates a high-ranking status of IDENT[sonorant] (and/or a markedness constraint against nasal geminates) as a recoverability condition.

Finally, in order to minimally distinguish the two root-final obstruents /t/ and /k/, with the former but not the latter assimilating to a heterorganic following voiceless obstruent, previous work such as Tateishi (1990), Padgett (1991 [1995]), and Ito and Mester (1996) has attributed the difference to featural underspecification, with /k/ specified as dorsal, but /t/ underspecified for place and acting as the default consonant (for crosslinguistic evidence for the choice of coronal as the default place, see Paradis and Prunet 1991 and references cited there). In OT terms, it is sufficient for the faithfulness constraint IDENT[dorsal] to be ranked higher than the alignment constraint, and IDENT[coronal] to be ranked lower (see Kurisu 2000 for statements of these IDENT constraints). Dorsality does not always trump coronality; this is rather subordinated to the onset-coda asymmetry. Thus the /t-k/ sequence in (17) – underlying /bet+koo/ – turns into *bekkoo* 'separate clause', but the /k-t/ case in (15) – underlying /hak+too/ – appears as *hakutoo* 'white sugar', not as **hakkoo*. Gemination is unidirectional, i.e., only from onset to coda, due to high-ranking IDENT-ONSET. Table (18) summarizes all changes at SJ compound junctures discussed in this section, with examples illustrating each case.

(18)

Rt-Final: / Rt-Initial:	V		n		k		t	
V	V.V	sai+ai 最愛	n.V	ren+ai 恋愛	*ku*.V	hak*u*+ai 博愛	*tu*.V	kat*u*+ai 割愛
r	V.r	kuu+ran 空欄	n.r	han+ran 反乱	*ku*.r	kyok*u*+ron 極論	*tu*.r	sot*u*-ron 卒論
w	V.w	gi+wak*u* 疑惑	n.w	kon+wak*u* 困惑	*ku*.w	tiku+wa 竹輪	*tu*.w	hat*u*+wa 発話
y	V.y	kyoo+yuu 共有	n.y	sin+yuu 親友	*ku*.y	gak*u*+yuu 学友	*tu*.y	kat*u*+yoo 活用
m	V.m	see+mon 正門	m.m	sem+mon 専門	*ku*.m	gak*u*+mon 学問	*tu*.m	sit*u*+mon 質問
n	V.n	zyoo+nai 場内	n.n	an+nai 案内	*ku*.n	kok*u*+nai 国内	*tu*.n	sit*u*+nai 室内
b	V.b	ee+bun 英文	m.b	rom+bun 論文	*ku*.b	sak*u*+bun 作文	*tu*.b	zat*u*-bun 雑文
z	V.z	kee+zai 経済	n.z	kan+zen 完全	*ku*.z	tyok*u*+zen 直前	*tu*.z	tot*u*+zen 突然
d	V.d	tee+den 停電	n.d	sen+den 宣伝	*ku*.d	syuk*u*+dai 宿題	*tu*.d	zet*u*+dai 絶大
g	V.g	dai+gak*u* 大学	ŋ.g	siŋ+gak*u* 進学	*ku*.g	dok*u*+gak*u* 独学	*tu*.g	tet*u*+gak*u* 哲学
p/h	V.h	koo+hai 後輩	m.p	sem+pai 先輩	*ku*.h	taku+hai 宅配	p.p	zip+pi 実費
s	V.s	ee+see 衛星	n.s	en+see 遠征	*ku*.s	gak*u*+see 学生	s.s	zis+sen 実践
t	V.t	kai+ten 開店	n.t	kan+tan 簡単	*ku*.t	dak*u*+ten 濁点	t.t	zit+tai 実態
k	V.k	koo+kan 交換	ŋ.k	buŋ+ken 文献	k.k	gak+koo 学校	k.k	zik.kan 実感

- Glosses for V-final column: *most beloved, empty column, doubt, joint ownership, front gate, on premises, English sentence, economy, power outage, university, one's junior, satellite, store opening, exchange.*
- Glosses for n-final column: *love, revolt, bewilderment, close friend, specialization, guide, thesis, perfect, advertisement, continuing education, one's senior, expedition, simple, reference.*

- Glosses for k-final column*: philanthropy, extreme argument, tube roll (of fish paste), school friend, study, domestic composition, immediately before, homework, self-taught, home delivery, student, turbid point, school.*
- Glosses for t-final column: *reluctantly delete, graduation thesis, utterance, conjugation, question, indoor, miscellaneous writings, sudden, enormous, philosophy, actual expense, actual practice, actual situation, realization.*

3.2 Word compounding and crisp edges

The previous sections outlined the segmental and prosodic properties of SJ roots and root compounds (i.e., instances of word$_{SJ}$ = root$_{SJ}$+root$_{SJ}$). At a higher level of SJ word structure, we find a rich system of word compounds (i.e., compound$_{SJ}$ = word$_{SJ}$#word$_{SJ}$). A first set of examples are what is commonly referred to as *yoji jukugo* 四字熟語, or 'four-character compounds', non-compositional idiomatic expressions as in (19), often borrowed from Classical Chinese and listed as such in dictionaries.

(19) Idiomatic four-character compounds
   a. zyaku+niku # kyoo+syoku 弱肉強食
      'weak+meat#strong+eat, the law of the jungle, the great fish eat the small'
   b. i+ku#doo+on 異口同音
      'different+mouths#same+sound, with one voice, unanimously'
   c. ee+ko#see+sui 栄枯盛衰
      'blossom+wilt#prosperous+decline, ups and downs (of life), rise and fall'

Besides their non-compositionality, such idiomatic four-character compounds form two phonological (accentual) phrases, e.g., {i+ku}{doo+on} etc., and are not subject to the compound junctural accent rule found in the productive word compounds to be discussed below.[9]

Setting aside such idiomatic compounds, word compounding is highly productive, and we find many compounds made up of two SJ words (themselves compounds at the root level), such as [koo+koo]$_{SJ}$#[ya+kyuu]$_{SJ}$ 高校野球 'high school baseball', [syuu+syoku]$_{SJ}$ #[si+en]$_{SJ}$ 就職支援 'employment assistance', etc.[10] Different from root compounding, word compounding is not restricted to members of the

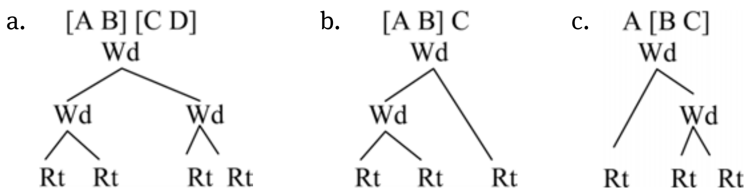
9 For details, see Ito and Mester (Ch. 9, this volume) on word formation and phonological processes, Kawahara (Ch. 11, this volume), and Kubozono, Ito, and Mester (1997).

10 Here and below, square brackets [. . .] indicate the constituent structure of the word, following standard practice.

SJ vocabulary stratum. Alongside the usual SJ#SJ combinations, we find other combinations, such as SJ#YAMATO and FOREIGN#SJ (see Ito and Mester 2003: 143 for examples illustrating all possible combinations): $[\text{ki+setu}]_{\text{SJ}}\#[\text{hazure}]_{\text{Yamato}}$ 季節はずれ 'off-season', $[\text{ken+kyuu}]_{\text{SJ}}\#[\text{fooramu}]_{\text{Foreign}}$ 研究フォーラム 'research forum', etc.

It is also possible to add a non-compounded single root$_{\text{SJ}}$ to a word$_{\text{SJ}}$ either on the right or the left, as in (20b,c). These have the appearances of 'suffixes' and 'prefixes', but since they also occur as compound root members, there is no clear and compelling morphological reason to differentiate between those occurring in the innermost word compound and those adjoined outside.

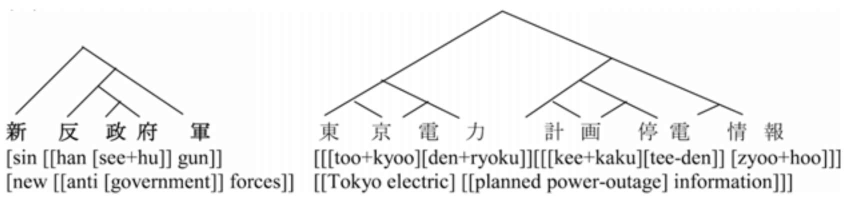
(20) Morphological structure of complex compounds



a.	b.	c.
[koo+koo] [ya+kyuu]	[kin+yuu] tyoo	sin [hatu+mee]
高校野球	金融庁	新発明
'high+school base+ball'	'financial agency'	'new discovery'
[syuu+syoku] [si+en]	[boo+ryoku] dan	bee [see+hu]
就職支援	暴力団	米政府
'employment assistance'	'gangster group'	'U.S. government'
[nen+kin] [mon+dai]	[tan+ken] tai	han [see+ki]
年金問題	探検隊	半世紀
'annuity problem'	'expedition team'	'half century'
[ren+sai] [syoo+setu]	[kai+ran] ban	betu [sya+kai]
連載小説	回覧板	別社会
'serialized novel'	'circulation notice'	'separate society'

Thus, besides the doubly-branching four-member compound structure [A B] [C D] in (20a), there are two kinds of three-member compounds: the left-branching structure [A B] C in (20b), and the right-branching structure A [B C] in (20c). Still more complex cases as in (21) can be reduced to these elementary configurations.

(21)



These long compounds act syntactically as single lexical items, hence a recursive morphological structure is appropriate.[11] McCawley's (1968) careful study of complex compounds shows a pattern distinctly different from that predicted by the generalizations obtained so far from simple two-member SJ root compounds. As shown in (22), with relevant word compound junctures indicated by #, a sequence of SJ roots such as /bet/ and /tee/ in (22a) surfaces with epenthesis in the complex compound across the #-juncture (/…bet*u*#tee…/), where our analysis so far predicts not disaligning epenthesis, but alignment-preserving gemination (/…bet#tee…/). As also shown in (22), the predicted lack of epenthesis continues to be correct at root compound junctures (/bet+tee/, etc.).

(22) Epenthesis vs. no epenthesis

	Epenthesis	No Epenthesis
a.	[AB][CD]	[BC]
	[toku+betu]#[tee+en] 特別庭園	[bet+tee] 別庭
	*[toku+bet]#[tee-en]	*[betu+tee]
	'special garden'	'annex garden'
b.	[AB]C	[BC]
	[toku+betu]#seki 特別席	[bes+seki] 別席
	*[toku+bes]#seki	*[betu+seki]
	'special seat'	'different seat'
c.	A[BC]	[AB]
	betu#[koo+moku] 別項目	[bek+koo] 別項
	*bek#[koo+moku]	*[betu+koo]
	'separate item'	'different reference'

Table (23) shows in a similar way that the alignment-wise expected p-allophone does not appear following a #-juncture in complex compounds.

(23) h/p alternation

	h-allophone	p-allophone
a.	[AB][CD]	cf. [BC]
	[kan+zen]#[hai+boku] 完全敗北	[zem+pai] 全敗
	*[kan+zem]#[pai+boku]	*[zen+hai]
	'total defeat'	'all defeat'
b.	[AB]C	cf. [BC]
	[man+nen]#hitu 万年筆	[em+pitu] 鉛筆
	*[man+nem]#pitu	*[en+hitu]
	'10000-year pen, fountain pen'	'lead-pen'; 'pencil'

---

11 Longer compounds are accentually different and have some characteristics of phrases, see Kubozono, Ito, and Mester (1997) for the description and Ito and Mester (2007) for the analysis of so-called long and overlong compounds.

c. A[BC] | cf. [AB]
sin#[hatu+mee] 新発明 | [ham+patu] 反発
*sim#[patu+mee] | *[han+hatu]
'new invention' | 'opposite+start'; 'rebel'

betu#[hyoo+ki] 別表記 | [bep+pyoo] 別表
*bep#[pyoo+ki] | *[betu-hyoo]
'separate transcription' | 'alternate list'

McCawley (1968), making imaginative use of the segmental boundary-based framework of SPE, developed an analysis whose essential insights are easily recaptured in the syntax-prosody interface theory (see Selkirk 2011, Ito and Mester 2013, and references cited there; see also Ishihara, this volume). In section 2, we stated the size restriction on SJ roots (1) (repeated in (24)) as demanding one foot.

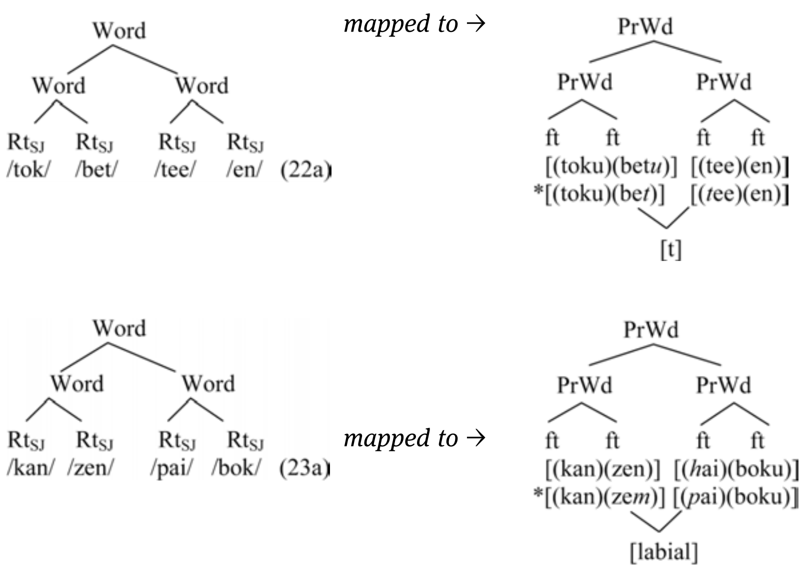
(24) Root size: | root_{SJ} | = ft (where ft ≤ 2μ, i.e., $\sigma_\mu\sigma_\mu$, $\sigma_{\mu\mu}$, or σ_μ)

Seen in a larger context, (24) is the manifestation of an interface constraint (25a) matching a morphological entity (here root_{SJ}) to a phonological constituent (here, a moraic-trochaic foot).

(25) Interface constraints
a. MATCH-ROOT$_{SJ}$-TO-ft
b. MATCH-LXWD-TO-PRWD

In tandem with the general interface constraint (25b) matching lexical words to prosodic words, (25a) maps the morphological structure (26a) to the prosodic structure in (26b) (parentheses indicate footing).

(26) a. Morphological structure b. Prosodic structure



Given this prosodic structure, it becomes apparent that the unexpected instances of epenthesis (instead of expected gemination) occur at the end of a prosodic word (PrWd), and the unexpected h-allophones at the beginning of a PrWd. What is avoided, then, is feature linkage across a prosodic word boundary, formally a consequence of the CRISPEDGE family of constraints (Ito and Mester 1999) requiring edges of prosodic constituents to be feature-wise closed. Crispness means that the constituent does not share features with (or is in other ways dependent on the properties of) adjacent prosodic constituents. The crucial member of the constraint family is CRISPEDGE(PRWD) requiring all PrWds, including the embedded ones in (26), to have crisp edges. In both [tokubet*u*][teeen] and [kanzen][*h*aiboku], every PrWd member has crisp edges, but in *[tokube*t*][*t*eeen] and *[kanze*m*][*p*aiboku], the geminate *tt* and the assimilated nasal-obstruent cluster *mp*, a partial geminate structure, share place features and violate the CRISPEDGE(PRWD) constraint. Without shared place features, CODACOND forces the epenthetic *u* in [tokubet*u*], and the *h*-allophone emerges initially in [*h*aiboku]. We interpret the nasal place assimilation sometimes found in the pronunciation of examples like [simbu*ŋ*][kookoku] 'newspaper advertisement' as a phonetic/fast speech phenomenon, as shown by its optionality and the fact that the genuinely phonological process resulting in /p/ for /h/ is not possible in this context (cf. [simbu*n*][*h*aitatu] 'newspaper delivery', *[simbu*m*][*p*aitatu], see Ito and Mester (1996: 38–39) and Kadono (2009) for discussion.

Geminate and partial geminate structures also violate CRISPEDGE(ft) and CRISPEDGE(σ). Both are low-ranking in Japanese, so simple PrWds like [(be*t*)(*t*ee)] and [(ze*m*)(*p*ai)] satisfy CRISPEDGE(PRWD) but violate the CRISPEDGE constraints for the smaller prosodic constituents. On the other hand, we saw in section 4 that ALIGN-ROOT-R (10) forestalls epenthesis at the right edge and creates geminate configurations whenever possible. This means that in terms of constraint ranking, ALIGN-ROOT-R is dominated by CRISPEDGE(PRWD), but in turn dominates both CRISPEDGE (ft) and CRISPEDGE(σ) (27).

(27) CRISPEDGE(PRWD) » ALIGN-ROOT-R » CRISPEDGE(ft), CRISPEDGE(σ)

The patterns of epenthesis and of the *h/p* alternation in complex compounds are thus straightforward consequences of (i) the mapping to prosodic structure through the interface constraints (25) and (ii) the interaction of the constraints governing this prosodic structure (27).

It is essential to conceive of the size restriction on root$_{SJ}$ in terms of a prosodic constituent – a single foot – and not in terms of raw mora counting. Even monomoraic SJ roots (like /si/ 詩 'poem' or /ku/ 句 'phrase') always constitute a foot, albeit a subminimal one. In OT terms, the FOOT BINARITY Constraint is outranked by the interface constraint demanding root$_{SJ}$ be matched with a foot (MATCH-ROOT$_{SJ}$ -TO-ft » FTBIN).

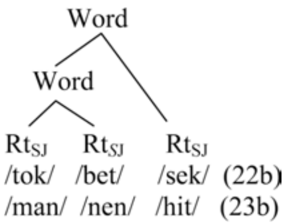
Once compounding has combined two such monomoraic roots into a PrWd, even though the absolute mora count is only 2, the word behaves as a two-foot struc-

ture $[\mu]_F+[\mu]_F$: CrispEdge(PrWd) ensures featural closure, and no geminate structure arises when further compounding takes place (see Kubozono 1993 for an independent argument from accentuation for the same conclusion). Thus, /ti+ku/ 地区 'locale' from monomoraic /ti/ 地 'land' and /ku/ 区 'division' is a prosodic word [(ti)(ku)] composed of two monomoraic feet, and as such is protected by CrispEdge (PrWd). This is why /bet#ti+ku/ 別地区 'different locale' is *betutiku* and not **bettiku*, and /bet#ki+ki/ 別機器 'different device' is *betukiki* and not **bekkiki*. The finding here reaffirms a fundamental tenet of prosodic phonology, viz., that the computation of prosody proceeds in terms of the constituents and categories of the prosodic hierarchy, and not in terms of an accounting system based on direct syllable or mora count (or some other unit measuring "weight" or "length").

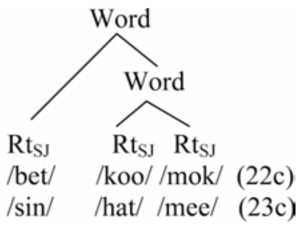
When we turn to three-member complex compounds, we find that another prosodic constraint may be at work, as argued in Ito and Mester (1996), namely, one militating against non-homogeneity of prosodic sisters. Applying the interface constraints (25) to the three-member compounds in (28) can in principle yield two kinds of prosodic structures, the adjunction structure in (29) and the homogeneous structure in (30). In (29), a right-adjoined foot (29a) and a left-adjoined foot (29b) are directly dominated by the highest PrWd. In (30), on the other hand, the lone foot projects a non-branching PrWd node, so that the immediate daughters of the highest PrWd are both PrWds, where the lone feet are type-lifted to be PrWds by themselves. Formally, the choice of structures would depend on the relative ranking of some prosodic structure minimization constraint (e.g., NoRecursivity or NoStructure) and a general prosodic homogeneity principle militating against adjunction structures (see also Myberg's 2010 constraint NoAdjunction, and Selkirk and Elordieta's 2010 analysis of Japanese and Basque phrasal phonology).

(28) Morphological structures

a. Left-branching

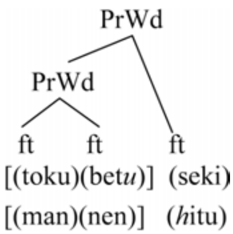


b. Right-branching

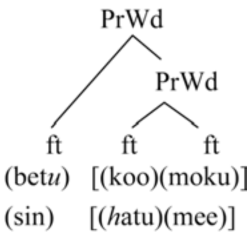


(29) Adjoined prosodic structures

a.

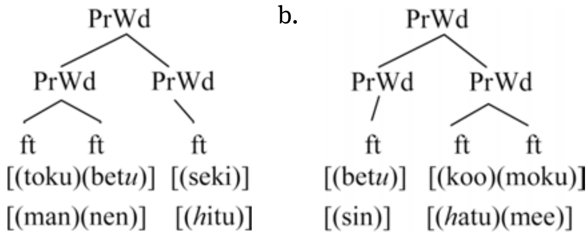


b.



(30) Homogeneous prosodic structures

a. b.



In the homogeneous prosodic structures (30), CRISPEDGE(PRWD) is directly responsible for all prosodic word edge effects: the epenthetic *u* in [[tokubet*u*]. . .] and [[bet*u*]. . .], and the *h*-allophone in [. . .[*h*atumee]] and [. . .[*h*itu]]. In the prosodic structures with adjunction (29), CRISPEDGE(PRWD) also directly ensures epenthesis at the end of the embedded PrWd ([[tokubet*u*]. . .] (29a) and the h-allophone at its beginning ([. . . [*h*atumee]] (29b)). The adjoined foot, on the other hand, even though itself not directly under the auspices of CRISPEDGE(PRWD), still cannot be linked to the adjacent featurally closed PrWd, and epenthesis is found in [(bet*u*) [. . .]] (29b), and the h-allophone in [[. . .](*h*itu)] (29a) as well. Lone adjoined feet and embedded PrWds thus pattern alike,[12] and either the adjunction structure or the homogeneous structure can explain the segmental alternations of SJ word compounding.

There seems to be no argument so far for the extra step of type-lifting to PrWd for reasons of homogeneity, contrary to Ito and Mester's (1996) earlier conclusions. However, looking beyond segmental alternations into the accentual arena, it turns out that the regular compound accentuation rules of Japanese apply in a way that argues for treating the lone foot as a type-lifted full-fledged PrWd, i.e., for the homogeneous structure (30) and against the adjunction structure (29). As is well known (see Kawahara, Ch. 11, this volume, and Kubozono 2008), in a compound word of the form $[[_{\mathrm{PrWd1}}\ldots][_{\mathrm{PrWd2}}\ldots]]$, compound accent is assigned at the juncture, namely, at the beginning of PrWd_2 (e.g., *nama-ta'mago* 'raw egg', *denki-ka'misori* 'electric razor'), except if PrWd_2 consists of only one foot, in which case, disregarding some complications, it is assigned to the end of PrWd_1 (*temuzu'-gawa* 'Thames river', *kamera'-man* 'camera man'). This generalization is fully obeyed for $\mathrm{word}_{\mathrm{SJ}}$ com-

12 See McCawley (1968: 116–118) for issues of optionality and sporadic counterexamples to this generalization. In a similar vein, Vance (1987: 161–162) gives examples such as zis#[sya+kai] 'the real world' as well as the well-known [san+kak]#kee '[three-angle] shape, triangle'. More detailed phonetic investigation along the lines of Beckman (1996) would be welcome to tease apart the relative roles in these cases of genuine vowel~zero alternation, a phonological process, and high vowel devoicing, a phonetic process, see Fujimoto (this volume). It will be important for any such investigation to take into account the fact that there are no counterexamples with /p/: zitu#[hee+ryoku], not *zip#[pee+ryoku] 'real [soldier strength], effective strength' but: zip-pi 'actual expenses', nor are there counterexamples in the middle of four-root combinations of the form [AB][CD] (i.e., between B and C) /[san+kak]#[kan+kee]/ 'triangular relationship, love triangle' is /[sankaku][kankee]/, never *[sankak][kankee].

pounds, so [[AB][CD]] structures receive accent on the initial syllable of the second member ([tokubetu][te'een] (22a), [kanzen][ha'iboku] (23a). Crucially, three-member compounds behave as if they are composed of two PrWd compound members: [[A][BC]] structures receive the accent on the initial syllable of the second member ([betu][ko'omoku]] (22c), [[sin][ha'tumee]] (23c)), and [[AB][C]] structures on the last syllable of the first member ([(toku)(betu'**)**] [(**s**eki)] (22b), [(man)(ne'n)][hitu]] (23b)), since the second member is (necessarily) one foot.

4 Concluding remarks

This chapter has outlined the basic characteristics of SJ roots, and their compounding behavior at the levels of the root and of the word. The alternations characteristic of SJ items are seen to be due (i) to the basic syllable structure constraints of Japanese and (ii) to alignment principles governing the mapping between morphosyntactic structure and prosodic structure. With a proper understanding of the prosodic structures involved and of the principles governing them, the phonology of SJ items follows from a few rather natural basic assumptions.

The specific property that sets SJ roots apart from the rest of the Japanese lexicon is their one-foot size limit, which in turn manifests itself in terms of alignment and interface constraints. Although they constitute only a small subset of the Japanese lexicon, the phonological study of SJ roots and compounds provides a revealing window into many aspects of Japanese phonology as well as phonological structure in general, including their segmental (epenthesis, gemination, nasal assimilation) and prosodic aspects (edge effects, structural effects, accentual implications, etc.).

References

Beckman, Jill N. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14. 1–46.

Beckman, Mary E. 1996. When is a syllable not a syllable? In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 95–123. Berlin, New York: Mouton de Gruyter.

Brame, Michael. 1974. The cycle in phonology: Stress in Palestinian, Maltese and Spanish. *Linguistic Inquiry* 5. 39–60.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73. 493–533.

de Lacy, Paul. 2002. *The formal expression of markedness*. Amherst, MA: University of Massachusetts dissertation. [ROA–542].

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology.* Cambridge: Cambridge University Press.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Gelbart, Ben and Shigeto Kawahara. 2007. Lexical cues to foreignness in Japanese. Paper presented at the Formal Approaches to Japanese Linguistics 4, 2007. 49–60. Cambridge, MA: MITWPL.

Goldsmith, John. 1990. *Autosegmental and metrical phonology.* Oxford and Cambridge, MA: Blackwell.

Inkelas, Sharon. 1989. *Prosodic constituency in the lexicon*. Stanford, CA: Stanford University dissertation. (Garland Press, 1990).

Ishihara, Shinichiro. this volume. Chapter 14: Syntax-phonology interface. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko. 1986. *Syllable theory in prosodic phonology.* Amherst, MA: University of Massachusetts dissertation. (Garland Press, 1988).

Ito, Junko and Armin Mester. 1993. Licensed segments and safe paths. In Carole Paradis and Darlene LaCharité (eds.), *Constraint-based theories in multilinear phonology*, 197–213. Toronto: University of Toronto Press.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Cambridge, MA: Blackwell.

Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko and Armin Mester. 1999. Realignment. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *Proceedings of the Utrecht Workshop on Prosodic Morphology*, 188–217. Cambridge: Cambridge University Press.

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure* (MIT Press Linguistic Inquiry Monograph Series 41). Cambridge, MA: MIT press.

Ito, Junko and Armin Mester. 2007. Prosodic adjunction in Japanese compounds. In Yoichi Miyamoto and Masao Ochi (eds.), *Formal approaches to Japanese linguistics: Proceedings of FAJL 4*, 97–111. Cambridge, MA: MIT Department of Linguistics and Philosophy.

Ito, Junko and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124. 20–40.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kadono, Jiro. 2009. A sonority-based approach to Sino-Japanese compounds. *Journal of Japanese Linguistics* 25. 75–111.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In In-Seok Yang (ed.), *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kubozono, Haruo. 1993. *The organization of Japanese prosody.* Tokyo: Kurosio.

Kubozono, Haruo. 2008. Japanese accent. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 165–191. Oxford: Oxford University Press.

Kubozono, Haruo. this volume. Introduction. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kubozono, Haruo, Junko Ito and Armin Mester. 1997. On'inkōzō kara mita go to ku no kyōkai: Fukugō meishi akusento no bunseki [The phonological boundary between the word and the phrase: An analysis of Japanese compound noun accentuation]. In Spoken Language Working Group (ed.), *Bunpō to onsei* [Speech and grammar], 147–166. Tokyo: Kuroshio.

Kurisu, Kazutaka. 2000. Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9. 147–85.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17. 267–302.

Martin, Samuel E. 1952. *Morphophonemics of standard colloquial Japanese*. Baltimore: Linguistic Society of America.

McCarthy, John J. and Alan Prince. 1993. *Prosodic morphology I: Constraint interaction and satisfaction*. Unpublished ms., University of Massachusetts, Amherst, and Rutgers University, New Brunswick, NJ.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Moreton, Elliot and Shigeaki Amano. 1999. The effect of lexical stratum phonotactics on the perception of Japanese vowel length. Paper presented at the 6th European Conference on Speech Communication and Technology (EUROSPEECH'99), Budapest, Hungary, 2679–2682.

Moreton, Elliot, Shigeaki Amano and Tadahisa Kondo. 1998. Statistical phonotactics in Japanese. Paper presented at the Acoustical Society of Japan, 63–70.

Myberg, Sara. 2010. *The intonational phonology of Stockholm Swedish*. Stockholm: Stockholm University dissertation. (Acta Universitatis Stockholmiensis, 2010).

Nasu, Akio. this volume. Chapter 6: Phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Padgett, Jaye. 1991 [1995]. *Stricture in feature geometry*. Amherst, MA: University of Massachusetts dissertation. (Revised version, CSLI Press, 1995).

Paradis, Carole and Jean-François Prunet (eds.). 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.

Poser, William J. 1988. Glide formation and compensatory lengthening in Japanese. *Linguistic Inquiry* 19. 494–502.

Prince, Alan S. and Paul Smolensky. 1993 [2004]. *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University and University of Colorado, Boulder Ms. RuCCS–TR–2. (Blackwell, 2004).

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology* 3. 371–405.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle and Alan Yu (eds.), *The handbook of phonological theory*, 2nd edition, 435–85. Oxford: Blackwell.

Selkirk, Elisabeth and Gorka Elordieta. 2010. *The role for prosodic markedness constraints in phonological phrase formation in two pitch accent languages*. Handout of paper presented at Tone and Intonation in Europe (TIE) 4, Stockholm University, Department of Scandinavian Languages. (http://www.hum.su.se/pub/jsp/polopoly.jsp?d=13236&a=73666, retrieved 11/29/2011.).

Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13. 209–235. Amherst, Massachusetts: GLSA.
Vance, Timothy J. 1987. *An introduction to Japanese phonology*. New York: SUNY Press.
Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge, U.K.: Cambridge University Press.
Vance, Timothy J. 2010. Renjō wa ikiteiru ka [Is *renjō* (liaison) a live process in Modern Japanese?]. Paper presented at the 5th Phonology Festa, Lake Biwako, Japan.

Haruo Kubozono

8 Loanword phonology

1 Introduction

Linguistic research generally tends to undervalue the study of loanwords as opposed to that of native words. Unfortunately, this is true in Japanese phonology, too. Most serious phonological analyses of loanwords are, in fact, relatively recent ones such as Quackenbush and Ohso (1990), Kubozono (1996, 2006a), Katayama (1998), Mutsukawa (2009) and Irwin (2011), although there are some earlier ones, such as Lovins (1975).

It is certainly true that loanwords display somewhat strange features that are not typically shown by native words. In contemporary Japanese, for example, [ti] and [tsa] are now acceptable sound sequences in loanwords, e.g., [ti:] 'tea', [mo:tsarɯto] 'Mozart', but not in native and Sino-Japanese (henceforth 'SJ') vocabulary: see Pintér (this volume) and Vance and Matsugu (2008) for more details about new sound sequences in modern Japanese; see also Nasu (this volume) for the lexical strata in the language. Similarly, most loanwords are lexically 'accented', i.e., involve a sudden pitch fall, whereas a majority of native words are 'unaccented' (Sibata 1994; Kubozono 2006a; see also section 7 below and Kawahara, this volume). These facts may be taken as suggesting that loanword phonology is substantially different from the phonology of native words, which, in turn, may imply that a study of loanwords does not shed new light on the structure of the host language.

In this chapter, we will challenge this idea by looking at Japanese loanword phonology from cross-linguistic perspectives. We will see that loanwords serve as a mirror that reflects the core structure of the host language that might not otherwise show itself clearly. To see this point, this chapter describes various aspects of loanword phonology in modern Japanese with focus on the two fundamental and interrelated questions of loanword phonology in general: (i) where loanword phonology comes from, and (ii) what it tells us about the structure of the host language, i.e., Japanese. The processes and phenomena to be discussed include vowel epenthesis, consonant epenthesis and glide formation, the asymmetry between /ai/ and /au/, syllable weight, accent, and truncation. These processes reveal the basic nature of Japanese phonology. For example, an examination of the loanword accent rule reveals that it is a rule of accented native words in general and resembles the Latin accent rule (section 7 below and Kawahara, this volume). The most important point is that the structure of Japanese loanwords reflects the core structure of the language which, in turn, is largely governed by language universal principles. Japanese is no different from other languages in this respect. Since 84% of loanwords in modern

Japanese originate from English (Sibata 1994), most phenomena to be discussed in this chapter are those observed when English words are borrowed into Japanese.

This chapter is organized as follows. The next section (section 2) discusses the ways Japanese borrows vowels and consonants from English and other foreign languages. As for vowels, we examine how umlaut, long vowels and schwas are borrowed in the language. The analysis of consonants addresses two related questions: (i) how the language copes with those consonant-vowel sequences that violate its traditional phonotactic constraints, and (ii) how new consonant-vowel sequences have been established. The second question is discussed in detail by Pintér (this volume). Section 3 addresses the question of vowel epenthesis in loanwords and specifically examines the strategies that Japanese employs to choose epenthetic vowels. Section 4 examines the phenomenon whereby glides are inserted to break up vowel sequences that would otherwise form a hiatus, i.e., vowel-vowel sequences across a syllable boundary.

Section 5 points out some crucial differences between /ai/ and /au/, which exhibit contrastive but consistent behaviors in several independent processes. Section 6 reviews past works on syllable weight, showing that various independent phenomena including consonant gemination and antigemination are constrained by a constraint prohibiting superheavy, i.e., trimoraic, syllables. Section 7 analyzes various phenomena concerning loanword accent and discusses, in particular, where the accent patterns of simplex loanwords and alphabetic acronyms come from and how the unaccented pattern comes about in these words. The final section (section 8) provides a summary of the chapter as well as some questions that remain for future work.

2 Segmental correspondences

2.1 Short vowels and schwas

By way of introduction to the loanword phonology of modern Japanese, let us examine how vowels and consonants have been borrowed. Modern standard Japanese has five short vowels and chooses one of these vowels for any short or lax vowel in the source language. To show the correspondences between tense vowels in English and long vowels in Japanese, we adopt the traditional transcriptions of English vowels in standard British English, or Received Pronunciation (RP) in this chapter (Jones 1960): e.g., [iː] ‘peak’, [i] ‘pick’, [ei] ‘taste’, [e] ‘test’, [uː] ‘pool’, [u] ‘pull’.

The basic rule used in loanword adaption is to choose the vowel that is phonetically closest to the source vowel. For example, English [æ] is usually borrowed as /a/ in Japanese. Since the latter language has fewer short vowels than the former, one and the same vowel in the latter often corresponds to more than one vowel in the former. For example, /a/ is chosen for three vowels in English as illustrated in

(1), of which the third pattern is observed only in some words (Lovins 1975; Quackenbush and Ohso 1990: 51–61). In (1) and the rest of this chapter, phonemic representations in / / are predominantly used for Japanese vowels and consonants, whereas phonetic representations in [ ] are used whenever they are more appropriate and/or the phonemic status of the sound is ambiguous.

(1) a. [æ] → /a/ bat.to 'bat', bak.ku 'back'
b. [ʌ] → /a/ bat.to 'but', kat.to 'cut
c. [ɔ] → /a/ sak.kaa 'soccer', ka.ku.te.ru 'cocktail'

Table 1 gives a summary of English lax vowels and the Japanese vowels that basically correspond to them (see Irwin 2011 for a more comprehensive discussion). It is worth adding here that English [æ] after velar stops, i.e., [k] and [g], tends to enter Japanese with the palatal glide /j/: e.g., /kjat.to/ 'cat', /kja.ra.me.ru/ 'caramel', /kjan.pu/ 'camp', /gjan.gu/ 'gang', /gja.gu/ 'gag' (Quackenbush and Ohso 1990: 56; Irwin 2011: 97). In these words, [æ] is "unpacked" into /j/ and /a/ in Japanese.

Table 1: Basic correspondences between lax English vowels and their Japanese counterparts

English vowel	Japanese vowel	Examples
[i]	/i/	pin, kiss, milk
[e]	/e/	pen, test, best
[ʌ]	/a/	but, cut, mother
[æ]	/a/	bat, back, lag
[u]	/u/	push, bull, put
[ɔ]	/o/	box, pot, dog

The fate of the English schwa is a little more complicated. English schwas and other reduced vowels in non-final position tend to be borrowed in different forms in Japanese depending on the spelling in the source language (Quackenbush and Ohso 1990: 86–95). This is shown in (2), where " " indicates the English spelling, and / / indicates the Japanese phonemic rendition. The fact that /a/ rather than /u/ is chosen for [ə] spelt with "u" suggests that /a/ may be the default vowel in Japanese for non-final schwas in English (and for final schwas, too, as we shall see in the next section).

(2) a. "a" → /a/ Japan, paradise, about, woman, camera, banana, Canada
b. "e" → /e/ camera, accent, Kennedy, system, elegant, garden, model
c. "o" → /o/ melody, parody, colony, balcony, iron, police, gorilla tomato
d. "i" → /i/ animal, cardigan, stamina, delicate, personality, justice, victim
e. "u" → /a/ campus, circus, curriculum, asparagus, focus, minus, suspense

2.2 Long vowels and diphthongs

In principle, long or tense vowels in the source language are borrowed as long vowels in Japanese. Since modern Japanese has only three diphthongs, /ai/, /oi/ and /ui/ (see Kubozono Ch. 5, this volume, and section 5 below), [ei] and [ou] in English generally enter Japanese as long vowels, that is, as /ee/ and /oo/, respectively.[1] English [ɔ:] is also borrowed as /oo/: Thus, 'coat'/'court', 'low, row'/'law, raw' and 'pose'/'pause' become homophones, respectively, in Japanese, as shown in (3). [iə], in contrast, enters the language as a heterosyllabic vowel sequence [i.a], as in /ni.a/ 'near': Syllable boundaries are denoted by dots (.) wherever necessary. Table 2 summarizes the correspondences between tense vowels and diphthongs in English and their counterparts in Japanese.

(3) a. [ou] → /oo/ kooto 'coat', roo 'low', poozu 'pose'

b. [ɔ:] → /oo/ kooto 'court', roo 'law', poozu 'pause'

Table 2: Correspondences between tense vowels and diphthongs in English and their Japanese counterparts

English vowel	Japanese vowel	Examples
[i:]	/ii/	meat, leak, team
[u:]	/uu/	pool, moon, boots
[ei]	/ee/	bacon, take, tape
[ou]	/oo/	coat, boat, toast
[ɔ:]	/oo/	court, short, talk
[ɑ:]	/aa/	car, star, market
[ə:]	/aa/	first, turn, Berkeley
[ai]	/ai/	bike, cider, rider
[au]	/a(.)u/	out, about, mouse
[ɔi]	/oi/	oil, coin, boil
[iə]	/i.a/	gear, near, rear
[ɛə]	/e.a/	air, rare, bear
[uə]	/u.a/	poor, tour, lure

While tense vowels and diphthongs in English are generally translated into long vowels or diphthongs, some undergo shortening in certain phonological contexts. For example, tense vowels as well as diphthongs in English undergo so-called "pre-nasal vowel shortening" by which they are shortened before the moraic nasal (Lovins 1975; Kubozono 1999a: see section 4 for some exceptions). This can be

1 Word-final [ei] in English monosyllables tends to be adopted as a diphthong in Japanese: [gei] 'gay' vs. [ge:mɯ] 'game'; [mei] 'May' vs. [me:do] 'maid, made', [me:kɯ] 'make-up'; [kei] 'the letter 'k'' vs. [ke:ki] 'cake', [ke:sɯ] 'case'; [rei] 'ray, lei' vs. [re:to] 'rate', [re:sɯ] 'race', [bɯre:ki] 'brake'; [dei] 'day' vs. [de:to] 'date'.

interpreted as a process preventing trimoraic syllables from occurring in the host language (see sections 5 and 6 below for full discussion).

(4) a. [aun] → /an/
fan.dee.syon, *faun.dee.syon 'foundation'
wan.dan, *wan.daun 'one down (in baseball)'

b. [ein] → /en/
su.ten.re.su, *su.tein.re.su 'stainless'
ken.bu.riz.zi, *keen.bu.riz.zi 'Cambridge'

c. [ɔ:n] → /on/
kon.bii.hu, *koon.bii.hu 'corned beef'

d. [i:n] → /in/
ku.in.bii, *ku.iin.bii 'queen bee'
gu.rin.pii.su, *gu.riin.pii.su 'green peas'

Another process involving a change in vowel quantity in loanwords turns English [au] to [a] before [ə] (Katayama 1998). As discussed in section 5 below, English [ai] does not shorten in loanwords to [a] in the same context (e.g., /tai.ja/ 'tire', /bai.jaa/ 'buyer'), showing an asymmetry with /au/ (see also Kubozono Ch. 5, this volume).

(5) a.waa, *au.waa 'hour'
hu.ra.waa, *hu.rau.waa 'flower'
pa.waa, *pau.waa 'power'

With the two major exceptions just mentioned, there is a high degree of correspondence between lax/tense distinctions in English and the short/long distinctions in Japanese.[2] There are two additional points to note here, both of which may suggest an influence of orthography (spelling) on pronunciations. First, English [ou] before the consonant cluster [st] tends to turn into long /oo/ ([o:]) in Japanese if they are spelt with two letters such as "oa", whereas they tend to be borrowed as /o/ if they are spelt with a single letter "o" in the source language. This can be seen from the data in Table 3. Thus, 'host' and 'post' are borrowed with a short vowel, whereas 'coast' and 'toast' have a long vowel in Japanese loanwords.[3]

2 There are some instances where English tense vowels turn into short vowels in Japanese for some unknown reason, e.g., /be.bii/ 'baby', /so.faa/ 'sofa' (vs. /sooda/ 'soda'), /redii/ 'lady', /mezjaa/ 'major', /raberu/ 'label', where the underlined vowel indicates the tense vowel in the English word.
3 This does not necessarily mean that the adaptation of English [ou] in Japanese is always sensitive to orthography. There are many instances of English [ou] borrowed as a long vowel in Japanese even if it is spelt with a single letter in the donor language: e.g., /hoomu/ 'home', /tookun/ 'token', /sumooku/ 'smoke'.

Table 3: Borrowing patterns of English [oust] in Japanese

Borrowed form English spelling	/o/	/oo/
"o"	host, hostess, most, post, poster	ghost
"oa"	–	coast, coaster, roast, toast, toaster

Secondly, word-final schwas display somewhat similar patterns. Schwas in this position are generally borrowed as /a/ or /aa/, but the choice of vowel length depends more or less on how the vowel is spelt in the source language. If the spelling contains the letter "r" in final position, it is likely to result in a long vowel in Japanese, whereas it tends to result in a short vowel if the spelling does not contain "r". For example, 'roller' is borrowed as /rooraa/, whereas 'Laura' enters Japanese as /roora/. This is exemplified in Table 4, where English words ending in a diphthongal vowel sequence (e.g., 'poor' [puə]) are excluded.

Table 4: English word-final schwas and their pronunciations in Japanese

Borrowed form English spelling	/a/	/aa/
"V"	Laura, fanta, soda, panda, Pola, festa, China, Panama, banana, sonata, Santa, data, pizza, mama, koala, coda, alpha	sofa (/so.faa/)
"Vr"	slipper, spanner, poplar,	calendar, radar, polar, color, sailor, donor, writer, roller, lighter, sister, center, soccer, Peter, piper, letter, pitcher, batter, butter

This tendency can be interpreted in two ways depending on whether the words were borrowed from British English or American English. These two varieties of English differ in whether the post-vocalic /r/ is pronounced or not. In the standard variety of British English known as Received Pronunciation (RP), the post-vocalic /r/ is not pronounced, which means that 'Pola' and 'polar' are homophonous, i.e., [poulə]. In the standard variety of American English, in contrast, the same /r/ is pronounced so that 'Pola' and 'polar' are not homophones: [poulə] vs. [poulɚ]. Supposing that the words were borrowed from British English, the asymmetrical distribution in Table 4 can be taken as evidence that orthography affects the choice of a sound in loanwords; that is, two letters (Vr) in the spelling are interpreted as long vowels in Japanese loanwords. If the words were borrowed from American English, on the other hand, the same data suggest that [ɚ] is "unpacked" into [a] and vowel length in the

host language. In other words, it is an instance where one sound in the source language splits into two in the host language. This resembles the well-known case where nasal vowels are unpacked into a sequence of an oral vowel and a coda nasal in a number of languages (Paradis and Prunet 2000). Japanese showed the same phenomenon many centuries ago when it attempted to integrate nasal vowels from Portuguese: e.g., [pã] → [pan] 'bread'. It is also similar to the case where umlaut vowels turn into a sequence of a glide and a vowel in Japanese (see the next section) as well as the case where the velar nasal [ŋ] in English is realized as a sequence of [ŋ] and [g] in Japanese (see section 2.4.2 below).

In relation to word-final schwas, it is probably worth adding that word-final [i] also tends to turn into a long vowel in Japanese although this is not dependent on the spelling in the source language: e.g., /re.dii/ 'lady, ready', /kjan.dii/ 'candy', /rak.kii/ 'lucky', /su.nuu.pii/ 'snoopy', /hap.pii/ 'happy', /ri.rii/ 'lily', /po.nii/ 'pony', /nan.sii/ 'Nancy'.

2.3 Umlaut vowels

One of the most interesting questions in loanword phonology is how the host language copes with sounds or sound patterns that are not present in its native phonology. One such case in the discussion of vowels concerns so-called "umlaut" vowels, which refer to front rounded vowels, namely, those vowels that have [+round, -back] in feature representation: high vowel [y] (often transcribed as [ü]) and mid vowel [ø] (=[ö]). Since Japanese does not have these vowels, it is confronted with a difficult situation when it borrows words from German and French, where umlaut vowels are abundant. Interestingly, Japanese employs apparently different solutions for vowels with different heights.

First, the high vowels [y] and [y:] are usually borrowed as /ju/ and /juu/, respectively. In this case, the [-back] feature of the input vowel is preserved in the palatal glide /j/, whereas [+round] is preserved in the output vowel. In other words, one segment in the source language is "unpacked" into two in the host language. On the other hand, the mid vowel [ø(:)] generally turns into /e(e)/ in Japanese.[4] In this case, the output form respects the [-back] feature of the input faithfully while ignoring the [+round] feature. These two patterns are exemplified below. If a conformity should be pursued, [y] could have turned into /i/ in (6a) or [ø] could have been borrowed as /jo/ in (6b).

4 Irwin (2011: 99) gives a different adaptation pattern for words from French: [ø:] → [u:] as in [ʃarɯtorɯ:zɯ] 'Chartreuse'. Note that Japanese has a variant pronunciation with [jɯ:] for this particular word: i.e., [ʃarɯtorjɯ:zɯ].

(6) a. [y(:)] → /ju(u)/
/mjun.hen/ ‘München’ (German)
/tjuu.rih.hi/ ‘Zürich’ (German)
/tjuu.bin.gen/ ‘Tübingen’ (German)
/ri.kjuu.ru/ ‘liqueur’ (French)
/a.ban.tjuu.ru/ ‘aventure’ (French)

b. [ø(:)] → /e(e)/
/ren.to.gen/ ‘Röntgen’ (German)
/ke.run/ ‘Köln’ (German)
/gee.te/ ‘Goethe’ (German)

The asymmetry between [y] and [ø] is interesting and poses the question of why [y] does not change into /i/ or why [ø] does not turn into /jo/. These hypothetical output forms are perfectly well formed in Japanese, which makes the asymmetry appear particularly mysterious (see Dohlus 2005, 2010 and Irwin 2011 for more data and theoretical analyses).

2.4 Consonants

2.4.1 Onset consonants

Consonants display a more complicated picture than vowels as there are more consonantal sounds than vocalic ones both in English and Japanese. Table 5 shows the basic correspondences between English and Japanese consonants in the onset position of the syllable (see Irwin 2011: 95 for a more comprehensive list): the loanword forms in Japanese are transcribed with broad phonetic symbols in order to be neutral to the controversy over their phonemic status (see Pintér, this volume, for full discussion).

Just like vowels, consonants follow the basic rule whereby they find the phonetically closest consonants in the host language as their counterparts. Thus, Japanese /t/ corresponds to English /t/ although they differ slightly in their places of articulation: the former is a dental plosive involving the blade of the tongue, whereas the latter is produced at the alveolar region of the palate and the tip of the tongue. Likewise, English /f/ has the bilabial fricative [ɸ], an allophone of /h/, as its counterpart in Japanese, despite the fact that they are produced at slightly different places, i.e., labio-dental vs. bilabial.

Again, one finds some cases where one and the same consonant in Japanese corresponds to more than one consonant in English. For example, /l/ and /r/ in English turn into /r/ (often transcribed as [ɾ]) in Japanese. Likewise, /s/ and /θ/ are both borrowed as /s/. Consequently, ‘lice’/‘rice’ and ‘sink’/think’ become homophones

Table 5: Basic correspondence of onset consonants

English	Japanese	Examples
[p]	[p]	pin, pot
[t]	[t]	top, ten
[k]	[k]	cup, kangaroo
[b]	[b]	bat, ban, best
[v]	[b]	vat, van, vest
[d]	[d]	dog, deck
[g]	[g]	group, google
[f]	[ɸ]	full, foot
[θ]	[s]	thank, three
[s]	[s]	sun, sleep
[ʃ]	[ʃ]	shy, ship
[h]	[h]	hat, hot
[h]	[ç]	hit, hip
[h]	[ɸ]	hook, hoop
[ð]	[z]	the, that
[z]	[z]	zebra, zone
[tʃ]	[tʃ]	Charles, charming
[dʒ]	[dʒ]	jam, jet
[si]	[ʃi]	sink, sit
[θi]	[ʃi]	think, theta
[ʃi]	[ʃi]	ship, shit
[ti]	[tʃi]	tip, tick
[tʃi]	[tʃi]	chip, chicken
[m]	[m]	moon, mat
[n]	[n]	net, knock
[l]	[r] ([ɾ])	light, lady, lice
[r]	[r]	right, ready, rice

in the host language. The same is true of /si/ and /ʃi/ in English, which merge into [ʃi] in Japanese, where 'sit' and 'shit', for example, become homophonous.

It is probably worth adding here that Japanese is not sensitive to the allophonic or subphonemic differences in the source language. For example, the presence or absence of aspiration in English does not affect the choice of consonant in Japanese: [p^{h}ai] → /pai/ 'pie', [spai] → /su.pai/ 'spy'; [t^{h}en] → /ten/ 'ten', [sta:] → /su.taa/ 'star'. Likewise, both the clear 'l' [l] and the dark 'l' [ɫ] in the source language are both borrowed as /r/ in the host language: [lip] → /rip.pu/ 'lip', [piɫ] → /pi.ru/ 'pill'.

2.4.2 Coda consonants

Coda consonants in English are borrowed into Japanese in much the same way as the onset consonants described in Table 5. Thus, /p, t, k/ in English turn into /p, t, k/ in Japanese (Table 6).

Table 6: Correspondences of coda consonants

English	Japanese	Examples
[p]	[p]	cap, top
[t]	[t]	pot, mat
[k]	[k]	pack, duck
[b]	[b]	cab, pub
[v]	[b]	love, live
[d]	[d]	bed, god
[g]	[g]	bag, big
[f]	[ɸ]	tough, knife
[θ]	[s]	bath, mouth
[s]	[s]	bus, mouse
[ʃ]	[ʃ]	push, smash
[ð]	[z]	clothes, bathe
[z]	[z]	nose, pose
[tʃ]	[tʃ]	beach, peach
[dʒ]	[dʒ]	bridge, judge
[m]	[m]	ham, gum,
[m]	[m] (moraic nasal)	camp, hamburger
[n]	[n] (moraic nasal)	ten, tent, tender
[ŋ]	[ŋg]	gang, long
[l]	[r] ([ɾ])	pill, pool

One notable difference between onset and coda consonants is that the process of vowel epenthesis accompanies the latter. This is a very productive process in the loanword phonology of Japanese, by which closed syllables in the source language are turned into open syllables in the host language. Some examples are given in (7). The choice of epenthetic vowel is discussed in section 3 below.

(7) *bath* → ba.su
love → ra.bu

The process of vowel epenthesis is often accompanied by the process of consonant gemination, which turns coda consonants in the source language into geminate consonants in Japanese. This is exemplified in (8). This process is heavily constrained with respect to the type of the coda consonant itself as well as the phonological context in which it appears (see section 6 below and Kawagoe, this volume, for a full discussion).

(8) *cap* → kyap.pu, *kya.pu
pot → pot.to, *po.to

Another difference between onset and coda consonants concerns the fate of /r/. Onset /r/ in English is borrowed as /r/ in Japanese as shown in Table 5 above, but

coda /r/ is not borrowed in the same way. Rather, together with the preceding nuclear vowel, it is borrowed as a long vowel, as shown in (9a). Schwas plus /r/ in English usually turn into /aa/ in Japanese, irrespective of the spelling of the schwa, as in (9b); see also the examples in Table 4 above.

(9) a. kaa 'car', baa 'bar', paa 'par', su.taa 'star', kaa.do 'card'
b. su.ka.raa 'scholar', taan 'turn', bat.taa 'batter', moo.taa 'motor', see.raa 'sailor'

The coda nasals /m/ and /n/ in English also exhibit a difference from their onset counterparts. First, coda /n/ is borrowed as a moraic nasal (often symbolized as /N/ in the literature): e.g., *ten* 'ten', *kan* 'can', *ten.to* 'tent', *den.baa* 'Denver'. The moraic nasal is phonetically manifested as a nasal vowel in word-final position as well as in word-medial position followed by an onsetless syllable; it is otherwise homorganic with a following consonant in other word-medial positions (Kawakami 1977: 81–84). Coda /m/ in English shows a more complicated picture than the coda /n/. Word-final /m/ in English is usually borrowed as /m/, which functions as an onset of the following syllable with an epenthetic vowel /u/, as in (10a). On the other hand, coda /m/ in English non-final position is usually realized as a moraic nasal, as in (10b).[5] Table 7 gives a summary.

(10) a. to.mu 'Tom', sa.mu 'Sam', bo.to.mu 'bottom', tai.mu 'time'
b. kyan.pu 'camp', han.baa.gaa 'hamburger', kan.pa.nii 'company', ko.ron.bi.a 'Columbia'

Table 7: Correspondences of nasals according to the positions in English

	Word-initial	Word-medial	Word-final
/m/	/m/ e.g., mat.to 'mat'	/n/ (/N/) e.g., kyan.pu 'camp' han.baa.gaa 'hamburger'	/mu/ e.g., ha.mu 'ham'
/n/	/n/ e.g., net.to 'net'	/n/ (/N/) e.g., ten.to 'tent'	/n/ (/N/) e.g., ten 'ten'

Finally, it is worth explaining how the velar nasal [ŋ] in English is borrowed into Japanese. While the basic rule of borrowing consonants is to use a single consonant in Japanese for a single consonant in English, [ŋ] displays a somewhat exceptional behavior. In English, this consonant appears only in the coda position. When borrowed into Japanese, it is "unpacked" into two consonants, [ŋ] and [g], of which the first

5 An English /m.w/ sequence exhibits an exceptional behavior: Cromwell → /ku.ro.<u>mu</u>.we.ru/, */ku.ro<u>n</u>.we.ru/.

establishes itself as a moraic nasal and the latter is followed by an epenthetic vowel. Thus, English 'ring' [riŋ] turns into a bisyllabic word /rin.gu/ [riŋ.gɯ] in Japanese. This is analogous to other cases of unpacking in loanword phonology. For example, nasal vowels split into an oral vowel and a coda nasal in old Japanese loanwords from Portuguese, e.g., [pã] → /pan/ 'bread'. Similarly, word-final [ɚ] in English is unpacked into two elements, /a/ and vowel length, e.g., /taan/ 'turn', /moo.taa/ 'motor', under the scenario that the relevant English words were borrowed from American English (section 2.2). Furthermore, the umlaut vowel [y] is unpacked into a sequence of the palatal glide /j/ and the vowel /u/, e.g., /mjun.hen/ 'München', /ri.kjuu.ru/ 'liqueur' (section 2.3).

2.5 Phonotactic constraints

Another interesting issue regarding the adaptation of sounds in loanword phonology concerns the question of how the host language copes with sound sequences that are not permitted in its own grammar. English has many consonant-vowel sequences that are not permitted by the phonotactic rules in the native and SJ vocabulary of Japanese. This includes the following: [si], [zi], [ti], [tu], [di], [du], [fa], [fi], [tsa], [wi], [hwi].

Many of these sequences are beginning to establish themselves in Japanese phonology, as we will see in the next section. For example, English [fa] and [fi] are usually borrowed as [ɸa] and [ɸi] in modern Japanese, e.g., [ɸatto] 'fat' and [ɸitto] 'fit', despite the fact that these output forms involve illegal consonant-vowel sequences in the traditional phonology of the language (see Pintér, this volume, for full discussion). In the traditional loanword phonology, in contrast, phonotactically illegal consonant-vowel sequences were borrowed with some modifications to respect the traditional phonotactic rules of the host language. These modifications seem to fall into two types.

The first type preserves the value of either the vowel or the consonant as such and changes the value of the other element in accordance with the phonotactic rules. In most cases, it is the vowel rather than the consonant that is faithfully preserved in the output. Thus, [si] generally turns into [ʃi] in Japanese, consequently merging with the original [ʃi] in the input: e.g., 'sip' and 'ship' become homophonous and are realized as /ʃip.pu/ in Japanese.

In some cases, however, the value of the consonant is exceptionally respected at the expense of the vowel. For example, English [ti] and [di] show two variant patterns in Japanese loanwords in addition to the new sound sequences [ti] and [di]. [ti] and [di] respectively turned into [tʃi] and [dʒi] in many words, just as [si] turned into [ʃi], but they turned into [te] and [de] in some words (Quackenbush and Ohso 1990; Crawford 2007). A good example showing this variation is the English phrase 'digital dilemma', which contains [di] in the initial syllables of the two component words. Interestingly, [di] in 'digital' turns into [de], while [di] in 'dilemma' is borrowed as [dʒi]: [dedʒitarɯ] + [dʒiremma]. This raises the question of why

Table 8: Modification of illegal sequences in Japanese

English	Japanese	examples
[si]	[ʃi]	silk, six
[zi]	[dʒi]	zipper, zigzag
[ti]	[tʃi]	tip, ticktock
[ti]	[te]	stick
[ti:]	[tʃi:]	team, steam
[tu]	[tsɯ]	tour
[tu:]	[tsɯ:]	two, tool
[di]	[dʒi]	credit, dilemma, studio
[di]	[de]	digital, demerit, handicap (> /hande/)
[di:]	[dʒi:]	diesel
[ʃe]	[se]	milk-shake
[fa]	[ɸa]	fax, fight
[fi]	[ɸi] (~[ɸɯ.i])	fit, film
[tsa]	[tsa]~[tsɯ.a]	Mozart
[wi]	[ɯi]	whisky, winter,
[we]	[ɯe]	west, waist

[de] rather than [dʒi] was chosen in the first word. A possible explanation may be that a sequence of [dʒi] was avoided in the output: The word would have become [dʒidʒitarɯ] if the vowel in the first syllable of the input had been faithfully preserved in the output.

Not surprisingly, native speakers of Japanese show different adaptation patterns depending on their age. For example, elderly speakers tend to replace English [ti] and [di] with native sequences, [te]/[tʃi] and [de]/[dʒi], whereas younger speakers adapt the input sequences as such. Thus, [ti] and [di] in *tissue*, *disco*, *Disney*, and *building* show age-related variations: [tʃitʃʃɯ] or [tetʃʃɯ] (elderly) vs. [tiʃʃɯ] (young), [desɯko] vs. [disɯko], [dezɯni:] vs. [dizɯni:], [birɯdʒiŋgɯ] vs. [birɯdiŋgɯ].

A second way of coping with illegal consonant-vowel sequences is to insert an epenthetic vowel between the two sounds, thereby converting /CV/ into /CV.V/. This strategy is taken by a limited number of consonant-vowel sequences like [tsa], [fi], and [wi]. Thus, [tsa] turns into the bisyllabic sequence of [tsɯ] and [a] as in [mo:tsɯarɯto] (~[mo:tsarɯto]) 'Mozart'. [fi] was also realized in two syllables in the traditional loanword phonology, [ɸɯ.i] : e.g., [ɸɯirɯmɯ] 'film'. Likewise, [wi] in English is also often decomposed into two syllables in Japanese, i.e., [wɯ] + [i], the first syllable being consequently reduced to [ɯ] due to a phonotactic constraint banning /wu/: e.g., /u.i.su.kii/ [ɯisɯki:] 'whisky'. Note that the location of accent in [ɸɯi'rɯmɯ] and [ɯi'sɯki:] suggests that there is a syllable boundary between [ɯ] and [i] in these words (see sections 5 and 6 for a related issue).

The two strategies to cope with illegal sequences are summarized in Table 8. Again, the loanword forms in the output are transcribed with broad phonetic symbols

in order to be neutral to the controversy over their phonemic status, e.g., whether [ʃ] has established itself as a phoneme independent of [s] (see Pintér, this volume, for full discussion).

3 Vowel epenthesis

Having looked at how English vowels and consonants are adapted in Japanese, let us see how syllable structures change in the course of borrowing. Like many languages in the world, Japanese does not freely tolerate closed syllables and consonant clusters. This reflects universal principles by which closed syllables and consonant clusters are avoided as much as possible. They are expressed, respectively, by the constraints NOCODA and *COMPLEX in Optimality Theory (Prince and Smolensky 1993).

(11) a. NOCODA: syllables end in a vowel (i.e., avoid syllables that end in a consonant).

b. *COMPLEX: No more than one C or V may associate to any syllable position node.

Closed syllables, i.e., syllables ending in a consonant, are more marked than open syllables, i.e., those ending in a vowel. Similarly, consonant clusters, e.g., /st/, are more marked than single consonants, e.g., /s/ or /t/. Clusters consisting of three consonants are even more marked than those consisting of two consonants. In any case, existence of a marked structure presupposes the existence of an unmarked structure in a single phonological system. Thus, a language that permits closed syllables also permits open syllables; a language with consonant clusters also admits single consonants. This relationship, which Roman Jakobson called the 'Implicational Law' (Jakobson 1968), can be seen in language acquisition, too. Namely, children who can produce closed syllables and consonant clusters can also produce open syllables and single consonants, but not the other way around. Not surprisingly, children tend to avoid producing these marked structures even in a language like English where the adult language tolerates them. (12) gives some examples from English-speaking babies: (12a) turns closed syllables into open ones, whereas (12b) simplifies consonant clusters (Yavaş 1998).

(12) a. dog → [da], [dada]
bed → [bæ]
fish → [fi]
cat → [kaka]
milk → [mi]

b. bread → [bed]
free → [fi:]
street → [ti:t]
stop → [tɔp]
please → [pi:z]

Closed syllables and consonant clusters are avoided as much as possible in Japanese, too. As for closed syllables, Japanese has two types of consonants that can stand in the coda position: nasals and voiceless obstruents. Voiced obstruents tend to be avoided as much as possible in the coda in Japanese. Thus, voiced obstruents in the coda position in English words tend to become voiceless in Japanese: e.g., /bag.gu/~/bak.ku/ 'bag' (Nishimura 2003; Kawahara 2006, 2011; see also Kawagoe, this volume, section 6.2).[6] This tendency is not restricted to loanwords since the same devoicing has occurred in some native words. For example, the adverb /ta.da/ 'merely, only' turned into /tat.ta/ when geminated for emphasis. Here, single /d/ alternates with geminate /tt/, not with /dd/. Similarly, /h/ is geminated into /pp/ in the process of reduplication, e.g., /ha/ 'leave' ~ /hap.pa/ 'many leaves', while it alternates with /b/ when simply voiced by *rendaku* voicing in ordinary compounds, e.g., /tja-ba/ 'tea leaf'. These alternations reflect nothing but a universal constraint on coda consonants. Coda is an unprivileged position within a syllable where phonological contrasts tend to be neutralized (Beckman 1998). In the case under consideration, voicing contrasts are lost in this syllable position with the result that only [-voice], or the unmarked value of voice in obstruents, surfaces. Obstruent devoicing in the syllable coda is observed in a wide range of languages including German and the speech of English-speaking babies (e.g., [bak] 'bag') (Yavaş 1998: 140).

In Japanese, both nasals and voiceless obstruents in the coda constitute a timing unit or "mora" by themselves, and are thus called 'moraic nasals' and 'moraic obstruents', respectively. They are often represented as /N/ and /Q/ in the literature. Of these, only nasals can stand in the word-final position. That is, presence or absence of a coda nasal is contrastive in word-final position, but this is not the case with coda obstruents.[7] In many languages, coda consonants do not have their own place of articulation, reflecting the universal constraint known as the 'Coda Condition' (Ito 1986). This is exactly true in Japanese, where both coda nasals and obstruents are phonetically homorganic with the onset of the following syllable: e.g., /an.pu/ [am.pɯ] 'amplifier', /kan.to/ [kan.to] 'Kant', /tan.ku/ [taŋ.kɯ] 'tank'. Word-final nasals as well as word-medial coda nasals followed by an onsetless syllable have a somewhat neutral place of articulation and are often described as nasalized vowels, e.g., [ũ] (Kawakami 1977: 84). As coda obstruents can only appear

6 If this neutralization in voicing occurs, /bakku/ 'bag' and /bakku/ 'back' become indistinguishable from each other.

7 Unlike Tokyo Japanese, Kagoshima Japanese allows coda obstruents to occur in word-final position. In this dialect, the presence or absence of a coda obstruent is contrastive in this position, too: e.g., /tet/ 'iron' vs. /te/ 'hand', /kat/ 'persimmon' vs. /ka/ 'mosquito'.

before a (voiceless) obstruent across a syllable boundary, they invariably form a ‘geminate obstruent’ together with this following consonant in Tokyo Japanese: e.g., /kat.to/ [katto] ‘cut’.

Turning to tautosyllabic consonant clusters, the only type of consonant cluster that supposedly exists in modern Japanese is the so-called “yōon” or “palatalized consonants”, which appear only before the nuclear vowel (but not in the coda): e.g., /kja.ku/ ‘guest’, /mja.ku/ ‘pulse’, /kjoo/ ‘today’. Palatalized consonants were originally borrowed from Chinese, which naturally explains why they are dominantly found in SJ morphemes in modern Japanese. They are also found in loanwords: e.g., /kjuu.to/ ‘cute’, /kjat.to/ ‘cat’. These palatalized consonants appear before the three back vowels, /u/, /o/, and /a/, but not before the two non-back vowels, /i/ and /e/. This distributional fact suggests that /a/ is phonologically a back (or non-front) vowel and forms a natural class with /u/ and /o/ in Japanese. In contrast, the phonological structure of the palatalized consonants themselves may be a matter of dispute. They can be analyzed as a consonant-glide cluster in the onset, as a single palatalized (vs. non-palatalized) consonant, or as a glide that is attached to the following vowel rather than to the preceding consonant. It is not clear which analysis is phonologically most appropriate (see Choi 2000 for a similar problem in Korean phonology).

Returning to the main topic, loanwords in Japanese follow the basic structures of the syllable just described. They avoid creating closed syllables and consonant clusters. When faced with these marked structures, Japanese adopts the same strategy that is employed by many other languages: that is, it inserts a vowel in an appropriate place (see Hall 2011 for vowel epenthesis in other languages). This process of vowel epenthesis is described in (13), where English words are given in the input and epenthetic vowels are put in < >.

(13) a. *root, route* → ruu.t<o> [rɯ:to]
roots → ruu.t<u> [rɯ:tsɯ]

b. *star* → s<u>.taa [sɯta:]
sky → s<u>.kai [sɯkai]

c. *street* → s<u>.t<o>.rii.t<o> [sɯtori:to]

It must be emphasized here that vowel epenthesis occurs for two independent reasons. In (13a) vowels are inserted in order to avoid closed syllables. In (13b), in contrast, vowels are epenthesized to avoid consonant clusters. Thus, the first two epenthetic vowels and the last one in (13c) are inserted for different reasons although they all lead to the creation of CV syllables in the output. In languages that tolerate consonant clusters but not closed syllables, output structures such as the one in (14a) will be chosen as optimal. On the other hand, languages that admit closed syllables but not consonant clusters may well show an output structure as in (14b). Since Japanese does not tolerate either of these marked structures, it takes the pattern illustrated in (13c).

(14) a. *street* → strii.t<o>

b. *street* → s<u>.t<o>.riit

A question that has attracted serious attention in the literature of Japanese loanwords concerns the choice of epenthetic vowel (see Uffmann 2006 for epenthetic vowels in loanword phonology in general). This is largely predictable as shown in (15) and Table 9 (Lovins 1975; Quackenbush and Ohso 1990); see section 6 for a discussion of consonant gemination.

(15) a. /o/ is inserted after the dental stops, [t] and [d]: [mi:t<o>] 'meat', [ri:d<o>] 'lead'

b. /i/ is inserted after the palatoalveolar affricates, [tʃ] and [dʒ], as well as after [k] in some archaic words: [pi:tʃ<i>] 'peach', [b<ɯ>riddʒ<i>] 'bridge', [iŋk<i>] 'ink'

c. In all other contexts, /u/ is inserted: [mapp<ɯ>] 'map', [kjaʃʃ<ɯ>] 'cash'

Table 9: Summary of epenthetic vowels in Japanese loanwords

English coda consonant	Japanese output	Examples
[t]	[to]	pot, mat
[d]	[do]	bed, god
[tʃ]	[tʃi]	beach, peach
[dʒ]	[dʒi]	bridge, bleach
[k]	[ki]	cake, strike, steak
[k]	[kɯ]	pack, duck
[p]	[pɯ]	cap, top
[b]	[bɯ]	cab, pub
[v]	[bɯ]	love, live
[g]	[gɯ]	bag, big
[f]	[ɸɯ]	tough, knife
[θ]	[sɯ]	bath, booth
[s]	[sɯ]	bus, loss
[ʃ]	[ʃɯ]	push, smash
[ð]	[zɯ]	clothes, bathe
[z]	[zɯ]	nose, pose
[ts]	[tsɯ]	boots, pants
[dz]	[dzɯ]	kids, goods
[m]	[mɯ]	jam, dam
[ŋ]	[ŋgɯ]	gang, long
[l]	[rɯ]	pill, pool

In addition to the three epenthetic vowels in (15), /a/ and /o/ are chosen in a very restricted context. In loanwords from German and Dutch, the voiceless velar fricative [x] turns into [h] and is often geminated. The epenthetic vowel chosen in

these words is often /a/: e.g., /bahha/ ‘Bach (composer)’, /mahha/ ‘Mach’. To be more precise, [h] is transparent to the process of vowel epenthesis (Uffmann 2006) with the result that the vowel preceding this consonant is simply copied. This explains why /a/ is chosen in the above examples, while /o/ is chosen in /gohho/ ‘Gogh’ (Dutch painter).

In terms of distribution, /u/ (=[ɯ]) is no doubt the default epenthetic vowel in Japanese. This distributional fact can be explained from a perceptual viewpoint (Kubozono 1999b). First of all, /u/ is the weakest vowel in Japanese in the sense that it is phonetically the shortest vowel (Sagisaka and Tohkura 1984) and is most prone to vowel devoicing (see Fujimoto, this volume, for more facts about vowel devoicing). For this reason, inserting /u/ will make the output sound most similar to the input: the word /mi.r<u>.k<u>/ ‘milk’, for example, is perceptually more similar to the input /milk/ than other output candidates such as /mi.r<a>.k<o>/ and /mi.r<e>.k<i>/. In OT terms, this means that an output candidate with an epenthetic /u/ is more faithful to the input than any other candidate, while satisfying successfully the two syllable structure constraints in (11). That is, the choice of /u/ as an unmarked epenthetic vowel is a result of constraint interaction, i.e., interaction between the markedness (well-formedness) constraints on syllable structure in (11) and faithfulness constraints requiring the correspondence between the input (foreign sounds) and output (adapted forms).

Why then are /i/ and /o/ inserted as mentioned in (15a) and (15b), respectively? As for /i/, it is possible to rely on an articulatory/perceptual similarity between this palatal vowel and the preceding palatoalveolar affricates. This raises the interesting question of why the palatoalveolar fricative [ʃ] usually takes /u/ rather than /i/ (e.g., /kjas.sj<u>/ [kjaʃʃɯ] ‘cash’, /s<u>mas.sj<u>/ [sɯmaʃʃɯ] ‘smash’.[8] This may be attributed to the fact that [ʃ] in English syllable codas is produced with lip rounding, which adds an [u]-like quality to the consonant.

Equally interesting is the fact that [k] chooses /i/ as an epenthetic vowel in some words. Quackenbush and Ohso (1990) propose a kind of vowel harmony between the epenthetic vowel and the vowel in the preceding context. Namely, /i/ may be inserted if [k] is preceded by a front vowel: e.g., /keeki/ ‘cake’, /dekki/ ‘deck’ vs. /dok.ku/ ‘dock’, /buk.ku/ ‘book’, /bak.ku/ ‘back’.[9,10] In other words, the [-back]

8 Quackenbush and Ohso (1990: 36–37) give two old borrowings with an epenthetic /i/ after [ʃ] : [kjaʃʃi] ‘cash’, [daʃʃi] ‘dash’. These words are pronounced with an epenthetic /u/ in modern Japanese: [kjaʃʃɯ], [daʃʃɯ]. [buraʃi] ‘brush’, [saʃʃi] ‘sash’, and [kaʃimia] ‘cashmere’ are the few examples that still have an epenthetic [i] after [ʃ].

9 This additional evidence from loanwords reinforces the view that /a/ is a back (non-front) vowel in the phonological system of Japanese.

10 Interestingly, a very similar effect of vowel harmony is observed in SJ morphemes where /i/ rather than /u/ tends to be epenthesized after /k/ if this consonant is preceded by a front vowel: the SJ morpheme 益 ‘benefit’, for example, has two pronunciations, /eki/ and /jak<u>/. See Tateishi (1990) and Ito and Mester (1996) for more details.

feature of the underlying vowel is copied in the epenthetic vowel. This does not occur in every word, however, since many words choose /ɯ/ as an epenthetic vowel after a front vowel: e.g., /pii.k<u>/ 'peak', /kik.k<u>/ 'kick'. Concerning this, Quackenbush and Ohso (1990) note that the choice of /i/ after [k] is characteristic of some old borrowings. This view is supported by the existence of such minimal pairs as /s<u>.t<o>.rai.k<i>/ 'strike (labor action)' vs. /s<u>.t<o>.rai.k<u>/ 'strike (in baseball)' and /b<u>.ree.k<i>/ 'brake' vs. /b<u>.ree.k<u>/ 'break', in which the forms with <i> are older loans than those with <u>. It is also substantiated by such archaic forms as /in.k<i>/ 'ink', whose modern form is /in.k<u>/. However, it remains unclear why [k] chose /i/ rather than /u/ in old borrowings. This change from /k<i>/ to /k<u>/ may be attributable to a change in pronunciation in the original language, i.e., English, or it may reflect a perceptual change on the part of Japanese speakers. This is an interesting topic for future research.

The choice of /o/ in (15a) can be accounted for on perceptual grounds. Japanese has a native assimilatory rule that affricates [t] and [d] into [ts] and [dz] (or [dʒ]), respectively (see Kubozono's Introduction to this volume for details). This allophonic rule is described in (16).

(16) a. /t/ → [ts] / __/u/

b. /d/ → [dz] / __/u/

Because of this rule, /tu/ and /du/ would automatically be turned into [tsɯ] and [dzɯ]. These hypothetical adapted forms are perceptually quite distinct from the input [t] and [d] even with /u/ as an epenthetic vowel. On the other hand, [to] and [do] keep the original consonant while containing a somewhat marked epenthetic vowel. Here Japanese is faced with a dilemma by which it has to choose one of the two possible options: taking the unmarked epenthetic vowel /u/ or keeping the original stop consonant. These two options would result in [mi:tsɯ] and [mi:to] for 'meat', respectively. Faced with this dilemma, Japanese chose the second option with only some exceptions to be noted shortly below. This choice turns out to be a reasonable one because it has made it possible to distinguish between [t] and [ts], which are distinctive in English and other languages. Thus, the distinction between English [t] and [ts] as in 'root, route'/'roots' and 'sheet'/'sheets' is well preserved in Japanese as they require different epenthetic vowels as shown in (13a).

While the perceptual explanation plus an additional functional account sounds largely reasonable, several questions remain. One of the most interesting questions concerns the contrast between SJ and non-SJ loanwords with respect to the choice of epenthetic vowel after /t/. Since old Chinese had many closed syllables, old Japanese epenthesized a vowel to turn them into open syllables: i.e., /CVC/ turned into /CV.C<V>/. Many SJ morphemes contain an epenthetic vowel in modern Japanese for this reason, but interestingly, they only choose between /i/ and /u/ as an epenthetic vowel. This is illustrated in (17). What is of interest here is the fact that /t/ in

the original coda chose /u/ and not /o/, as illustrated in (17b) (Hayasi 1982; Ito and Mester 1996).

(17) a. tek → te.k<i> 'enemy'
sek → se.k<i> 'seat'

b. tet → te.t<u> [tetsɯ] 'iron'
bat → ba.t<u> [batsɯ] 'punishment'
kyak → kya.k<u> 'visitor'
rak → ra.k<u> 'comfort'

A historical account for this contrast between SJ morphemes and non-SJ borrowings may be that Japanese acquired the affrication rule in (16) after it had borrowed SJ morphemes from Chinese. Namely, Japanese tolerated [tɯ] and [dɯ] as a phonetic manifestation of /tu/ and /du/ when the words in (17) were adapted into Japanese. This historical account needs to be supported by other pieces of independent evidence, but seen conversely, this is suggestive of the possibility that loanword phonology provides an insight and a new perspective into a historical study of language.

It may be worth pointing out here that the rule in (15a) admits several exceptions. /u/ instead of /o/ is inserted after [t] in some words, typically in words where [t] is followed by another consonant in the source language. Thus, the English words 'tree' and 'twitter' are pronounced with an epenthetic /u/, i.e., [tsɯri:] and [tsɯitta:]. The word 'country' has two pronunciations, [kantori:] and [kantsɯri:]. 'Cutlet' also takes an epenthetic /u/, i.e., [katsɯretsɯ]. The choice of /u/ after /t/ is often found in relatively old borrowings from English, but it would be interesting to look for some (possibly phonetic) reasons for this group of exceptions.

4 Consonant epenthesis and glide formation

Our next topic concerns a phonological structure called 'hiatus' and the linguistic ways to resolve this marked structure. Hiatus refers to a sequence of vowels without any intervening consonant. It is widely known that this structure is disfavored by a number of languages including Japanese (Kindaichi 1976). In constraint-based accounts, this is due to the two constraints in (18). (18a) militates against diphthongs, or tautosyllabic vowel sequences, whereas (18b) bans vowel sequences across a syllable boundary.

(18) a. *COMPLEX: No more than one C or V may associate to any syllable position node; i.e., no complex vowel is allowed.

b. ONSET: Every syllable begins with a consonant. (i.e., no syllable begins with a vowel).

Apart from creating a diphthong out of this marked structure, there are generally four independent strategies to resolve it, as illustrated in (19) (Casali 1996/1998, 2011).

(19) a. Consonant insertion: VV → VCV
  b. Glide formation: VV → GV(:)
  c. Vowel elision: $V_1V_2 \rightarrow V_2$ (or V_1)
  d. Vowel coalescence: $V_1V_2 \rightarrow V_{12}$

Interestingly, all these four solutions are employed in Japanese, both as historical and contemporary synchronic processes (see Kubozono Ch. 5, this volume, for a full discussion of diphthongs and vowel coalescence). They are exemplified in (20), where < > and /+/ denote an epenthetic element and a morpheme boundary, respectively, while ( ) shows the origin of the word.[11]

(20) a. pi.a.no → /pi.<j>a.no/ 'piano' (loan)
  i.tari.a → /i.ta.ri.<j>a/ 'Italy' (loan)
  ko.a.ra → /ko.<w>a.ra/ 'koala bear' (loan)
  ha.ru+a.me → /ha.ru+<s>a.me/ 'spring rain' (native)
  mas + ao → /mas+<s>a.o/ 'pure blue' (native)

  b. bariumu → /ba.rjuu.mu/ 'barium' (loan)
  karusiumu → /ka.ru.sjuu.mu/ 'calsium' (loan)
  iu → /juu/ 'to say' (native)
  riu →/rjuu/ 'dragon' (SJ)

  c. a.ra+i.so → /a.ri.so/ 'rocky coast' (native)
  tai+iku → /tai.ku/ 'physical education, training' (SJ)

  d. naga+iki → /na.ge.ki/ 'long breath, lament' (native)
  sugoi → /su.gee/ 'wonderful' (native)

Of these four processes, the first two are productive in loanword phonology. This fact itself is very interesting and it is very important to ask why loanwords do not generally undergo the processes in (20c,d). Logically speaking, /sutoraiki/ 'strike' may well undergo the process in (20d), hence turning into /sutoreki/ or /sutore:ki/, but this is quite unlikely. This may be accidental, but it is also possible to relate this to the fact pointed out by Kubozono (1997) that input accent is quite well preserved

11 /nagaiki/ in (20d) turned into /nageki/ in old Japanese, where vowel length was not distinctive. In modern Japanese, where vowel length is distinctive, diphthongs generally turn into a long vowel as exemplified by the second word in (20d), /sugoi/. This is a case of compensatory lengthening that is generally observed in a quantity-sensitive prosodic system (see Kubozono Ch. 5, this volume).

in loanwords when they form the second member of compounds, i.e., that loanwords are more faithful to the input than native and SJ words with respect to the preservation of input accent in compound accentuation (see note 16). The fact that loanwords do not undergo the elision and coalescence processes in (20c,d) may be another instance showing that they are generally more faithful to the input than native and SJ words.

The two processes that actually occur in loanwords, i.e., (20a) and (20b), are both optional in Japanese phonology. (20a) thus creates a variation between [piano] and [pijano] for the English word /piano/. Like the glide formation in (20b), this process produces a glide that was not present in the input. Unlike (20b), it involves inserting either /j/ or /w/ in loanwords. The choice of the glide depends on the preceding vowel: /i/ and /e/ take /j/, whereas /u/ and /o/ choose /w/. Stated differently, the glide that is epenthesized is homorganic with the preceding vowel such that the palatal glide /j/ is inserted after a palatal (i.e., front) vowel, while the velar glide /w/ is inserted after a velar (i.e., back) vowel. This constitutes evidence for progressive place assimilation across a syllable boundary and can be compared with the regressive place assimilation known as Coda Condition, by which the coda consonant is assimilated to the onset consonant of the following syllable.

While (20a) involves both /j/ and /w/ in loanwords, the glide formation process in (20b) only involves creating the palatal glide /j/ in the output. The process of glide formation itself is very natural and found across languages, too. For example, [ius] 'use' and [riud] 'rude' turned into [ju:s] and [ru:d] in the history of English (Moore and Marckwardt 1981; Kubozono and Honma 2002). One should not overlook, however, that glide formation in (20b) does not involve producing the velar [w]. That is, while /iu/ turns into /ju/, /ui/ does not turn into /wi/ or /wi:/ (consequently into /i/ or /i:/): e.g., /uisukii/ 'whisky' → */wi.su.kii/, */i.su.kii/. This asymmetry between /iu/ and /ui/ is very interesting and may be linked to the general tendency whereby vowel sequences ending in /u/ turn very easily into monophthongs whereas those ending in /i/ are resistant to this process in Japanese in general (see the discussion in section 5 below and Kubozono Ch. 5, this volume, for details).

Another interesting aspect of the glide formation process in (20b) is that it is usually accompanied by the lengthening of the vowel following the glide (Poser 1988). Thus /bariumu/ turns into /barjuumu/ and not /barjumu/. This lengthening represents a very general process known as 'compensatory lengthening' by which the phonological weight or length of the input word/syllable is preserved in the output (Hayes 1989). This process is observed across languages as evidenced by the two English words cited above, i.e., 'use' and 'rude'. The fact that this lengthening occurs in Japanese loanwords provides additional evidence for the mora in Japanese as a unit of phonological weight (Kubozono 1999a).

Finally, the comparison between (20a) and (20b) raises a very interesting question. Given the two processes in (20a) and (20b), one can reasonably ask why the word /piano/ 'piano' was subject to (20a), while the word /bariumu/ underwent the

rule in (20b). Since the first vowel in the vowel sequences is /i/ in both cases, /piano/ could undergo the rule in (20b) and turn into [pja:no]. Similarly, /bariumu/ might well have undergone the rule in (20a), turning into [barijɯmɯ]. Obviously, these logical possibilities were ruled out by some principle or constraint. A descriptive generalization may simply be that /ia/ turns into [ija] and /iu/ into [jɯ:], but we can go one step further to ask why such a generalization is obtained. A more serious analysis may illuminate some new constraints on the two rules, which may then be collapsed into one rule.

5 Asymmetry between /au/ vs. /ai/

In her insightful work on loanword phonology of Japanese, Katayama (1998) argued that /ai/ and /au/ show different patterns. She pointed out that [ai] and [au] before a schwa [ə] in English are borrowed in different phonological forms. This is illustrated in (21a,b), where the English forms are given in the input.

(21) a. /taiə/ 'tire' → [tai.ja]
/faiə/ 'fire' → [ɸai.ja:]
/baiə/ 'buyer' → [bai.ja:]

b. /tauə/ 'tower' → [ta.wa:]
/sauə/ 'sour' → [sa.wa:]
/pauə/ 'power' → [pa.wa:]
/auə/ 'hour' → [a.wa:]
/flauə/ 'flower' → [ɸɯ.ra.wa:]

English /aiə/ turns into a bisyllabic form [ai.ja] in Japanese with the palatal glide [j] added as the onset of the second syllable. On the other hand, /auə/ undergoes the deletion of /u/ to yield the form [a.wa:] or, alternatively, /u/ is weakened to become the velar glide [w]. In this latter case, too, the resultant form is bisyllabic, with [w] functioning as the onset of the second syllable. However, the crucial difference between the two cases is evident. In the case of /aiə/, both /a/ and /i/ survive as a moraic element in the resultant loanword form, whereas /u/ loses its moraic status in the case of /auə/. Note that /au/ appears as freely as /ai/ in other phonological contexts, as exemplified in (22). However, it is clear that Japanese somehow avoids creating /au/ in the phonological context in (21). There is no comparable constraint on the occurrence of /ai/.

(22) <u>au</u>.to 'out'
r<u>au</u>.do 'loud'
p<u>au</u>.daa 'powder'

The asymmetry in (21) is interesting by itself, but a truly interesting point is that this represents a very general rule or constraint in Japanese phonology (Kubozono 2001a, 2005, 2008a). The asymmetry in question can be extended to native and SJ words, it can be generalized to cover vowel sequences other than /ai/ and /au/, and it is instrumental in explaining 'exceptions' in many other phenomena.

In the historical phonology of Japanese, for example, vowel sequences ending in /i/, i.e., /ai/, /ui/, /oi/ and /ei/, are much more stable than those ending in /u/, i.e., /au/, /ou/, /eu/ and /iu/. The vowel sequences in the first group have resisted the process of vowel coalescence, while those of the second group turned into monophthongs very easily.[12] Vowel coalescence does certainly occur in the first group, too (e.g., /daikon/ → /de:kon/ 'radish'), but it remains an optional process that occurs only in a certain speech style (in men's casual speech, to be exact) in only some dialects, including Tokyo Japanese. In contrast, coalescence of /Vu/ was obligatory in native and SJ words, and occurred independent of speech style and dialects. Some examples are given in (23).

(23) a. /au/ → /oo/
taka + u → takau → ta.koo 'high'
ahuta → auta → oo.ta 'meet (past)'
kau → koo 'high'
kyau → kyoo 'capital, home town'

b. /eu/ → /oo/
tefutefu → teuteu → tyoo.tyoo 'butterfly'

c. /iu/ → /juu/
iu → yuu 'to say'
riu → ryuu 'dragon'

The asymmetry between /ai/ and /au/ is observed not just in historical phonology, but accounts for a wide range of synchronic phenomena in modern Japanese. Japanese is subject to a constraint prohibiting superheavy, i.e., trimoraic, syllables, which is called the 'trimoraic syllable ban' (Kubozono 1999a). One consequence of this constraint is that long vowels and diphthongs are often shortened when they are followed by a coda nasal. Namely, /VVn/ is converted into /Vn/ by either shortening long vowels or deleting the second part of diphthongs (Kubozono 1995a, 1999a). Although this shortening/deletion process admits some exceptions, as we

12 There are more than ten patterns of vowel coalescence in Japanese, but all can be reduced to a simple rule whereby the resultant vowel inherits a [high] feature from the first vowel and other features from the second vowel (Kubozono Ch. 5, this volume). Interestingly, this is essentially the same as the coalescence rule found in many African languages (Casali 1996/1998).

will see shortly below, it occurs only in loanwords because trimoraic syllables were generally absent in native and SJ morphemes.

(24) a. English /aun/ → Japanese /an/
gu.ran.do 'ground'
fan.dee.syon 'foundation'
me.rii.goo.ran.do 'merry-go-round'
wan.dan 'one down' (in baseball)
tuu.dan 'two down' (in baseball)
wan.ban 'one bound (ground ball)' (in baseball)

b. English /eːn/ → Japanese /en/
ren.zi 'range'
tyen.zi 'change'
a.ren.zi 'arrange'
su.ten.re.su 'stainless'
en.zye.ru 'angel'
ken.bu.rid.dzi 'Cambridge'
men.te.nan.su 'maintenance'

c. English /iːn/ → Japanese /in/
gu.rin.pii.su 'green peas'
ma.sin 'machine'
ku.in.bii 'queen bee'

d. English /ɔːn/ → Japanese /on/
kon.bii.hu 'corned beef'
ron.rii 'lonely

The shortening process sketched in (24) is not a recent finding. Lovins (1975) described it over several decades ago and Kubozono (1995a) proposed to explain it in terms of a constraint on the maximal weight of the syllable. However, these previous studies apparently overlooked an interesting asymmetry between /ain/ and /aun/. Namely, there is no instance that involves shortening of /ain/ into /an/; /ain/ is invariably manifested as shown in (25).

(25) sain 'sign', rain 'line, The Rhine', rain.ga.wa 'River Rhine', de.zain 'design', ko.kain 'cocaine'

This strongly contrasts with the fact that /aun/ is shortened to /an/ in many instances including those in (24a). There are exceptions to (24a), as we shall see shortly below, but this does not negate the contrastive behavior of /ain/ and /aun/. In fact, /au/ patterns with long vowels and tends to become a short monophthong.

This means that the second element of /au/ behaves as if it were segmentally invisible when preceding a moraic nasal. This asymmetry between /ai/ and /au/ provides further evidence that /au/, but not /ai/, is unstable in modern Japanese.

The instability of /au/ as against /ai/ is also observed in compound truncation. The most productive pattern of this morphological process in contemporary Japanese is to form a four-mora word by combining the initial two moras of one component word with those of the other (Ito 1990; Ito and Mester 1995; Kubozono 1999a, 2003a; Ito and Mester, this volume). Some examples are given in (26), where { } denotes a bimoraic foot boundary.

(26) <u>se.ku</u>.sya.ru <u>ha.ra</u>.su.men.to → {se.ku}{ha.ra} 'sexual harassment'
<u>po.ke</u>t.to <u>mon</u>.su.taa → {po.ke}{mon} 'Pokémon, pocket monster'
<u>han</u>.gaa <u>su.to</u>.rai.ki → {han}{su.to} 'hunger strike'
<u>waa</u>.do <u>pu.ro</u>.ses.saa → {waa}{pu.ro} 'word processor'

This default pattern, however, admits several types of exceptions, one of which concerns /aun/ sequences. As suggested above, there are quite a few exceptions to the shortening process in (24a). Some are given in (27), where syllable boundaries are not specified because of potential ambiguity.

(27) s<u>aun</u>do 'sound', m<u>aun</u>ten 'mountain', k<u>aun</u>siru 'council', k<u>aun</u>to 'count'

The rule sketched in (26) predicts that the words in (27) will preserve the initial two moras in this morphological process: e.g., /saundo/ → /sau/, /maunten/ → /mau/. The fact is, however, that the moraic nasal is retained instead of the second half of /au/. This pattern is obtained whether /aun/ appears in the first component (28a) or in the second component (28b) (cf. Kuwamoto 1998b).

(28) a. <u>saun</u>do torakku → {<u>san</u>}{tora}, *{sau}{tora} 'sound track'

b. buruu <u>maun</u>ten → {buru}{<u>man</u>}, *{buru}{mau} 'Blue Mountain'
buritissyu <u>kaun</u>siru → {buri}{<u>kan</u>}, *{buri}{kau} 'British Council'
noo <u>kaun</u>to → {noo}{<u>kan</u>}, *{noo}{kau} 'no count (in baseball)'

In contrast, /ain/ and /oin/ do not show any such irregularity. There are not many truncated compounds that involve /ain/ or /oin/, but those that do follow the regular pattern by preserving the initial two moras of the trimoraic sequences. This is exemplified in (29).

(29) a. donto m<u>ain</u>do → {don}{m<u>ai</u>}, *{don}{man} 'Don't mind'

b. zyo<u>in</u>to bentyaa → {zyo<u>i</u>}{ben}, *{zyon}{ben} 'joint venture (business)'

The contrast between (28) and (29) suggests that the second mora of /aun/, i.e., /u/, is invisible to the morphological rule of compound truncation. Interestingly, long vowels and geminate obstruents (or moraic obstruents) often show a similar effect of invisibility in the same morphological process (Kubozono 1999a, 2002, 2003a; Kuwamoto 1998a,b; Ito 2000). As mentioned above, /au/ and long vowels show the same behavior in pre-nasal vowel shortening, i.e., they omit their second component. It is indeed interesting that /au/ patterns with long vowels rather than with /ai/ in compound truncation, too.

The same type of asymmetry between /ai/ and /au/ shows up in accentuation. In Tokyo Japanese, compound accent (CA) usually falls on the final syllable of the first member if the second member is one or two moras long (Akinaga 1981; McCawley 1968; Kubozono 1997).[13] Interestingly, the two vowel sequences in question show different patterns in this accentuation. If the first member ends in /ai/, the CA usually falls on /a/, as illustrated in (30a). This suggests that /ai/ belongs to one and the same syllable and, hence, that it is a diphthong. On the other hand, if the first member ends in /au/, the CA docks on /u/ rather than /a/. This is exemplified in (30b). In (30) and subsequent examples, accent is denoted by an apostrophe (') and placed immediately after the vowel that bears the accent.[14]

(30) a. ma'sai + zo'ku → masa'i-zoku, ?masai'-zoku 'Masai, tribe; the Masais'

b. do'nau + kawa' → donau'-gawa, *dona'u-gawa 'The River Donau (Danube)'

Some speakers seem to place the CA on /i/ in (30a), but no speaker puts the CA on /a/ in (30b). This contrast suggests that /ai/ forms a unified syllable in loanwords, whereas /au/ constitutes two separate syllables. Not surprisingly, the same syllabification seems to hold in Kagoshima Japanese, a dialect spoken in the south of Japan which is syllable-based rather than mora-based (like Tokyo Japanese). In this dialect, loanwords are usually accented, i.e., bear a high tone, on the penultimate syllable. /ai/ and /au/ pattern differently with respect to this accent rule, as shown in (31) (Kubozono 2004a, 2007a; Kubozono Ch. 5, this volume): high-toned syllables are denoted by capital letters.

(31) a. MA.sai 'Masai (name of a tribe in Africa)'
PAI.ron 'Pairon (brand name of a medicine)'
NAI.ru 'the River Nile'

b. do.NA.u 'the River Donau, or Danube'
pa.U.ro 'St. Paul'
to.ra.U.ma 'trauma'

13 In the constraint-based analysis to be discussed in section 6.2 below, this means that a CA is placed on a non-final, rightmost (bimoraic) foot of compound (Kubozono 1995b, 1997, 2008b, 2011).
14 "?" means that the form is marginally acceptable.

In (31a), /ai/ is counted as one syllable, whereas a syllable boundary falls between /a/ and /u/ in (31b). Although this observation must be borne out by a quantitative study, it seems that /ai/ and /au/ pattern differently in both the mora-based system of Tokyo Japanese and the syllable-based system of Kagoshima Japanese. In terms of syllabic organization, this means that /ai/ readily forms a diphthong, whereas /au/ resists integration into one unified syllable across dialects. This is consistent with the observation mentioned above, that /ai/ but not /au/ has stability as a diphthong.

So far we have seen several independent pieces of evidence for an asymmetry between /ai/ and /au/, which is deeply rooted in Japanese phonology. It is important to emphasize that all the analyses along this line were initiated by phonological research on loanwords, particularly by the work of Katayama (1998). Since /au/ does not occur in native and SJ morphemes in modern Japanese, the insight into the asymmetry between /ai/ and /au/ can only be obtained through analysis of loanwords. Further study may reveal the asymmetry in question in a wider range of phenomena in Japanese. This reinforces one of the main claims of this chapter, that is, that loanwords provide a very important source of data for understanding the nature of the language itself.

With this insight, we can go one step further and ask if the asymmetry in question is only characteristic of Japanese or represents a language-universal property. Kubozono (2008a) points out a certain asymmetry between /ai/ and /au/ in the phonology of English, Korean, and Romanian, but it is desirable to look at a wider range of languages. If it turns out that the asymmetry is shared by many other languages in the world, we can ask why such an asymmetry emerges cross-linguistically.[15] This will potentially contribute to general phonology and phonological theory.

6 Syllable weight and consonant gemination

Another area in which analysis of loanwords has contributed greatly to the study of Japanese phonology concerns the notion of syllable weight. As mentioned in the preceding section, Japanese displays a strong tendency to avoid superheavy, i.e., trimoraic, syllables. It is well-known that this syllable type is disfavored in a wide range of languages such as Hausa (Hayes 1986), English and other Germanic languages (Árnason 1980), Koya and Fula (Sherer 1994), and Pali (Zec 1995), to mention

15 This seems fairly likely since the monophthong [u] is more marked than the monophthong [i] cross-linguistically. This idea can be supported statistically by the UCLA Phonological Segment Inventory Database (UPSID), which shows that most two-vowel systems in the world's languages consist of [a] and [i] rather than [a] and [u]. Moreover, it is also in accordance with Stevens's (1989) claim that the vowels [a] and [i] are the two most acoustically stable vowels, representing anchor points in the vocal tract.

just a few (see Hayes 1995: 303 for more languages). Since this marked syllable type does not generally occur in native and SJ morphemes (see note 17), traditional Japanese phonology has overlooked the fact that Japanese shares the tendency in question with many other languages in the world.

We have already seen in (24) that the process of "pre-nasal shortening" in Japanese is triggered by pressure to avoid trimoraic syllables. This process is just one manifestation of a rather general constraint on syllable structure known as the "trimoraic syllable ban". Trimoraic syllables are resolved in different ways in different languages. Typical solutions are given in (32).

(32) a. vowel shortening/deletion: VVC → VC
b. coda deletion: VVC → VV
c. resyllabification: VVC → V.VC

Interestingly, all these solutions are observed in Japanese, where they conspire to avoid creating a trimoraic syllable. (32a) has already been described in (24) above. (32b) and (32c) are illustrated in (33) and (34), respectively (see Kubozono 1995a and 1999a for a more detailed analysis). In (33), the moraic nasal /n/ is deleted after a bimoraic vowel sequence. This change yields a bimoraic syllable out of a sequence that would otherwise result in a trimoraic one.

(33) entertainment → /en.taa.tei.men.to/, ?/en.taa.tein.men.to/
alignment → /a.rai.men.to/, ?/a.rain.men.to/

(34) a. sa'in + ka'i → sai'n-kai 'autograph + party; autograph signing party'
ra'in + kawa' → rai'n-gawa 'Rhine + river; The River Rhine'
deza'in + ha'ku → dezai'n-haku 'design + exposition; The Design Exposition'
barenta'in + de'e → barentai'n-dee 'St. Valentine's Day'
supe'in + kaze → supei'n-kaze 'Spain + cold; Spain Flu'
ko'in + syo'o → koi'n-syoo 'coin + dealer; coin dealer'
b. guri'in + sya → gurii'n-sya, ?guriin'-sya 'green + car; first-class car of a train'
me'en + syu'u → mee'n-syuu, *me'en-syuu 'Maine + state; the State of Maine'

(34) shows the accentual behavior of what appears to be a trimoraic syllable: a diphthong-like sequence followed by a moraic nasal in (34a) and what looks like a long vowel followed by a moraic nasal in (34b). As illustrated in (30) above, compound nouns with a monomoraic or bimoraic second member tend to bear accent on

the final syllable of the first member.[16] In the compounds in (34), the CA usually falls on the second mora of the vowel sequences rather than their first mora. This fact suggests that there is a syllable boundary within the vowel sequences in question, namely, that the first and second moras in the vowel sequences belong to different syllables: e.g., /sa.i'n.kai/, /ko.i'n.sjoo/, /me.e'n.sjuu/. This constitutes evidence that /VVn/ is (re)syllabified into /V.Vn/ in avoidance of a single trimoraic syllable. Interestingly, the same result is obtained from an analysis of the syllable-based system of Kagoshima Japanese. In this dialect, /Vn/ is accented, i.e., high-toned, in such words as /saIN-kai/, /barentaIN-dee/, /koIN-sjoo/ and /meEN-sjuu/, which suggests that what appears to be a trimoraic syllable actually consists of two syllables (Kubozono 2004a).

Apart from the three processes in (32), Japanese exhibits one more clear case where trimoraic syllables are avoided. This process, known as antigemination, blocks the otherwise uniform process of gemination. Consonant gemination inserts a moraic obstruent, or *sokuon*, before a voiceless obstruent, creating a geminate consonant (see Kawagoe, this volume, and Kubozono 2007b, 2013b for full discussion of this phenomenon). It has the effect of creating a heavy syllable, or a heavy-light syllable sequences word-finally, as shown in (35).

(35) *cup* → kap.pu, *ka.pu
hit → hit.to, *hi.to
cut → kat.to, *ka.to

This process is blocked, however, if the nuclear vowel is complex, i.e., a long vowel or diphthong. This is illustrated in (36).

(36) *carp* → kaa.pu, *kaap.pu
heat → hii.to, *hiit.to
cart → kaa.to, *kaat.to

The similarity between (35) and (36) is obvious: In both cases a heavy syllable is created in the output. In (35), gemination has created a heavy syllable out of a light syllable, whereas in (36) antigemination has blocked the creation of a superheavy syllable in favor of a heavy syllable. Gemination and antigemination thus conspire to yield a heavy syllable in preference to a superheavy syllable. In the originally monosyllabic words in (35) and (36), these processes conspire, in conjunction with vowel epenthesis (discussed in section 3) to produce bisyllabic forms consisting of a heavy and a light syllable.

16 To be more precise, monomoric and bimoraic second members tend to preserve their lexical accents in the compound if they are not accented on the final syllable and if they are not SJ morphemes (see section 7.2 below and Kubozono 1997 for a detailed discussion)

It is important to add here that this particular bisyllabic form represents one of the most unmarked word structures in Japanese (Ito et al. 1996; Kubozono 2000, 2003a). Specifically, bisyllabic words consisting of two heavy syllables and those composed of a heavy plus light syllable are the most preferred word forms in various phenomena in the language. These phenomena include babies' production and perception of early words, formation of *zuuja-go* (or *zuuzya-go*, musicians' language), loanword truncation (Ito and Mester, this volume), intensification of mimetic expressions (Nasu, this volume), and sporadic instances of vowel shortening and lengthening. The common feature exhibited by all these phenomena can be properly understood if and only if one invokes the notion of syllable weight in Japanese phonology. Without this notion, heavy-light bisyllables cannot be properly distinguished from light-heavy bisyllables, the latter being the most disfavored prosodic form in Japanese (Ito et al. 1996; Kubozono 2000, 2003a).

All in all, the arguments presented in this section demonstrate the importance of introducing the notion of syllable weight into Japanese phonology. The three-way distinction in syllable weight – light (monomoraic), heavy (bimoraic) and superheavy (trimoraic) syllables – plays a pivotal role in explaining various phenomena that would otherwise remain unaccounted for. The notion of 'superheavy syllable' has been motivated by a phonological study of loanwords, while this type of syllable does not generally occur in native and SJ vocabulary.[17] Here, again, loanwords have provided a crucial insight into the ways in which phonological analyses are to be carried out.

7 Accent

7.1 Loanword accent rule

The notion of syllable weight also provides a significant insight into the nature of Japanese accent (for full discussion and overview of Japanese accent, see Kubozono 2008b, 2011 and 2013a as well as Kawahara, this volume). And in this analysis too, loanwords play a key role. Let us begin with the traditional accent rule for loanwords in Tokyo Japanese, which is given in (37) and exemplified in (38) (McCawley 1968).[18] Syllable boundaries are also mora boundaries, although not necessarily vice versa. Again, apostrophes denote word accent, or the position where an abrupt pitch drop occurs in phonetic outputs.

17 There are a small number of native words that seem to have a superheavy syllable. /to'otta/ 'passed (past tense of *pass*)' is one such word which can be compared with /hasi'tta/ 'ran (past tense of *run*). The fact that this word bears accent on its initial mora rather than its second mora suggests that /toot/ forms a trimoraic syllable.

18 See Shinohara (2000) for an Optimality-theoretic analysis of loanword accent in Japanese.

(37) Place an accent on the syllable containing the antepenultimate mora, i.e., the third mora from the end of the word.

(38) pu'.ra.su 'plus'
ha'.wai 'Hawaii'
ba'.na.na 'banana'
ka'.na.da 'Canada'
ku.ri.su'.ma.su 'Christmas'
wa.si'n.ton 'Washington'
ma.ku.do.na'.ru.do 'McDonald's'

One fundamental question that naturally occurs is where this rule comes from. A major difference observed in loanwords and the other two types of words (native and SJ) is that the former type of word is mostly accented, whereas the latter two types of words prefer the unaccented pattern, or the pattern where no abrupt pitch drop occurs. According to Kubozono (2006a,b), only 7% of trimoraic loanwords are unaccented, whereas a majority of native and SJ trimoraic words are unaccented. This observation led some phonologists to assume that loanword accentuation is basically different from that of native and SJ words (Sibata 1994; Shinohara 2000). As demonstrated by Kubozono (2006a), however, loanword accentuation is not greatly different from native and SJ accentuation if we focus on accented words. In fact, a majority of *accented* native and SJ words basically follow the rule in (37). Some examples are given below.

(39) a. native Japanese words
i'.no.ti 'life'
na.ga'.sa.ki 'Nagasaki'
a.o'.mo.ri 'Aomori (Prefecture)'
hi.ma'.wa.ri 'sunflower'
a.ka'.gai 'ark shell'

b. SJ words (compounds)
tyu'u.go.ku 'China'
ka'n.ko.ku 'Korea'
ga.ku'.mon 'learning'

The next question to ask then is why loanwords prefer the accented pattern to the unaccented pattern. Kubozono (2006a) attributes this to the fact that English words, which are by far the biggest source of loanwords in Japanese, are produced with a pitch fall when pronounced in isolation. For example, the English word 'Washington' is produced with a sudden pitch fall between the first and second syllables in citation form. Since a sudden pitch fall is the distinctive feature of Japanese

accent, native speakers of the language show sensitivity to this pitch change and process loanwords as 'accented' as opposed to 'unaccented'. This account is based on both perceptual and phonological factors in loanword prosody: how source words are perceived by the speakers of the recipient language is affected by L2 phonetics and is also constrained by L1 phonology.

One may naturally wonder at this point why Japanese accent patterns are often different from those of the source language, e.g., why the word 'Washington' is accented on the second syllable in Japanese despite the fact the source word is accented (stressed) on the initial. This leads us back to the loanword accent rule in (37), which originates from native prosody. This interpretation provides a principled account for the fact that one and the same word is pronounced in different ways in different dialects (Kubozono 2006a, 2010, 2011). One example is given below, where capital letters indicate high-pitched syllables/moras for the sake of description.

(40) Tokyo: ma.KU.DO.NA.ru.do 'McDonald'
Kyoto: ma.ku.do.NA.ru.do
Kagoshima: ma.ku.do.na.RU.do
Koshikijima (Kagoshima Prefecture): MA.KU.DO.na.RU.do

The dialectal differences in (40) and similar regional differences that many other loanwords exhibit in accent patterns reflect nothing but the differences in accent systems or, more crucially, differences in the rule for *accented* native words.

Having understood that the rule in (37) accounts for the basic accent pattern of accented (as opposed to unaccented) nouns in Tokyo Japanese in general, let us compare it with the famous Latin accent rule (Hayes 1995), which is given below with the accented syllables highlighted by capital letters.

(41) Accent the penultimate syllable if it is heavy; if it is light, accent the antepenultimate syllable.
e.g., for.TUU.na 'fortune'
a.lex.AN.der 'Alexander'
PO.pu.lus 'people'
IN.te.grum 'perfect'

Differences between the rule in (37) and that of (41) are obvious. (37) is basically a mora-based rule in which the basic location of accent is determined by a mora-counting procedure. In contrast, (41) is a syllable-based rule where phonological distance is measured primarily in terms of the syllable. While these two rules look quite different from each other as they stand, their basic similarity becomes evident once they are reinterpreted in terms of syllable weight. Assuming that syllables are either heavy (bimoraic) or light (monomoraic) in both Japanese and Latin, the two rules predict the following accent patterns for the eight logically possible combinations

of three syllables at the end of words. H and L stand for heavy and light syllables, respectively.

(42) (=37) a. . . .L'LL# b. . . .LH'L# c. . . .LL'H# d. . . .LH'H#
e. . . .H'LL# f. . . .HH'L# g. . . .HL'H# h. . . .HH'H#

(43) (=41) a. . . .L'LL# b. . . .LH'L# c. . . .<u>L'LH#</u> d. . . .LH'H#
e. . . .H'LL# f. . . .HH'L# g. . . .<u>H'LH#</u> h. . . .HH'H#

A comparison of (42) and (43) reveals the same accent patterns in six out of the eight contexts. In other words, the two rules have largely the same effects. Thus, an analysis invoking the notion of syllable weight enables us to understand the basic similarity between Japanese and Latin accentuation. Since the Latin rule is shared by English and many other languages in the world (Hayes 1995), it follows that as far as accented words are concerned, Japanese noun accentuation is basically the same as the accentuation of these languages (Kubozono 1999a,b, 2006a,b, 2008b).

In relation to this, it is important to note that the Japanese accent rule is changing in the direction of the Latin accentuation. That is, Tokyo Japanese is now undergoing an accent change in the environments given in (42c,g), namely, in the two contexts in which Japanese accentuation was different from Latin accentuation. According to Kubozono's (1996) statistical work, about 80% of loanwords with the syllable structures in (42c) and (42g) are now accented on the antepenultimate rather than the penultimate syllable. Some examples are given below.

(44) c. L'LH#
bi'.gi.naa 'beginner'
do'.ra.gon 'dragon'
re'.ba.non 'Lebanon'
ra'.ma.dan 'Ramadan'
ta'.ri.ban 'Taliban'
a'.ma.zon 'Amazon'

g. H'LH#
i'n.ta.byuu 'interview'
e'n.de.baa '(Space Shuttle) Endeavor',
myu'u.zi.syan 'musician'
o'o.di.syon 'audition'

The new accent patterns illustrated in (44) cannot be explained as a simple imitation of English pronunciations since some words like 'beginner', 'endeavor' and 'musician' are accented on the penultimate syllable in English (i.e., *begínner*, *endéavor*, *musícian*). This suggests that the accentual change is not based on the

original stress pattern of the individual words but is quite systematic in Japanese phonology.[19]

Note here that native and SJ words in Japanese do not follow the accent change illustrated in (44). Thus /a.ka'.gai/ in (39a) and /ga.ku'.mon/ in (39b) remain accented on the penultimate syllable – or, equivalently, on the syllable containing the third mora from the end of the word. However, all these words are morphologically complex, i.e., compounds, and thus their accentuation can be accounted for by the general compound accent rule to be discussed in the next section.

7.2 Loanword accent and compound accent

In traditional studies of Japanese accent, loanword accent and compound accent (CA) have been formulated as entirely different rules (McCawley 1968; Akinaga 1981). Loanwords follow the antepenultimate accent rule in (37) above, whereas CA is supposed to fall into two kinds depending on the phonological length of the second member of compounds. Against this traditional formulation, Kubozono (1995b, 1997) proposed a new formulation of CA in the framework of Optimality Theory. This analysis is capable of generalizing the two kinds of CA rules using a set of general principles (constraints). This has been made possible by integrating the syllable and foot as well as the mora into the analysis of Japanese accent. In descriptive terms, this new formulation consists of the following five basic principles (or constraints in OT terms) given in the order of importance. Terms in the parentheses denote constraints in the optimality-theoretic analysis.

(45) a. Avoid placing/preserving an accent on the last syllable (Nonfinality–head syllable).
b. Preserve the accent of the second member of compounds (Max-accent).
c. Avoid placing/preserving an accent on the word-final bimoraic foot (Nonfinality–head foot).
d. Put the accent at or near the boundary between the first and second member of compounds (Align-accent).
e. Place an accent towards the end of the word as much as possible (Edgemostness).

19 This accentual change cannot account for the accentuation of some words such as /a'.ku.sen.to/ 'accent', /a'.do.bai.su/ 'advice' and /a'.ku.se.sa.rii/ 'accessory'. It is not clear whether they are only exceptions to the general accent rule or their accentuation represents a systematic change of loanword accentuation. A statistical study is needed.

Now let us apply this formulation to loanwords. Since loanwords are basically monomorphemic in Japanese, we can ignore the constraints in (45b,d). The rest of the formulation in (45) predicts the following accent patterns for the eight syllable structures in (42). We assume here a minimally specified foot structure, but assuming a fully specified foot structure will not affect the main results of this analysis. Again, { } denotes a foot structure.

(46) a. . . .{L'L}L# b. . . .L{H'}L# c. . . .{LL'}H# or . . .{L'L}H#
d. . . .L{H'}H# e. . . .{H'}LL# f. . . .H{H'}L#
g. . . .H{L'}H# or . . .{H'}LH# h. . . .H{H'}H#

The predicted accent patterns in (46) are essentially identical to those given in (42). Note that foot structure becomes ambiguous in several environments, depending on whether unfooted syllables are permitted or, equivalently, whether monomoraic (i.e., degenerate) feet are tolerated. However, this does not make any crucial difference in accent assignment except in (46g). In this particular environment, accent will dock onto the penultimate light syllable if monomoraic feet are permitted, i.e., . . .H{L'}H#, whereas it will fall on the antepenultimate heavy syllable if monomoraic feet are banned, i.e., . . .{H'}LH#. The latter pattern represents the new accent pattern shown in (44g). Moreover, different accent loci are predicted for (46c) depending on whether accent is permitted to fall on the right-hand syllable in {LL} feet. This will yield a variation between {LL'}H# and {L'L}H#, the latter being the new and more dominant accent pattern illustrated in (44c).

Despite these variations, the fact still remains that the CA rule formulated in (45) makes basically the same predictions as the accent rule in (37). If this is the case, it will follow that the loanword accent rule in (37) can be totally dispensed with. What we need is the compound accent rule outlined in (45), which should now be understood as a general, unified accent rule for Japanese nouns, both simplex and compound (see Sato 2002 and Kubozono 2004b for additional evidence for treating simplex loanwords in the same way as compound nouns). In this analysis, simplex and compound nouns are subject to one and the same accent rule, but can yield different patterns depending on their morphological complexity. In OT terms, compound nouns are subject to the five constraints in (45), whereas simplex words are exempt from the constraints in (45b,d) because of their morphological non-complexity.

Seen in this light, the accentual difference between loanwords in (44) and native/SJ words in (39) can also be accounted for in a reasonable way. Since all SJ morphemes and most native morphemes are one or two moras long, three-mora or longer native and SJ words are usually morphologically complex and involve a morpheme boundary. In these complex words, the principle/constraint in (45d) exerts its effect and, in combination with other constraints in (45), yields the results

in (47a). In contrast, loanwords are basically monomorphemic and will display the pattern in (47b).

(47) a. (=39) {a.ka'}+gai, {ga.ku'}+mon
b. (=44c) {a'.ma}zon, {ra'.ma}dan

In sum, the discussion in this section reveals that simplex and complex nouns in Japanese follow one and the same accent rule (see Kawahara, this volume, for more details). The only difference between these two types of nouns is that compounds, but not simplex words, involve a morpheme boundary and, hence, are subject to constraints that specifically concern morphologically complex words. The accentual differences between loanwords and native/SJ words also naturally follow from a difference in morphological complexity.

7.3 Unaccented loanwords

It was mentioned above that native and SJ words are strikingly different from loanwords in that many are unaccented. One mystery in Japanese phonology has been how the lexical distinction between accented and unaccented words is determined. An analysis of loanword accentuation provides a certain insight into this mystery. Here, again, the notion of syllable weight plays a pivotal role.

It is well known that, in Tokyo Japanese, only ten percent of loanwords are unaccented, whereas a majority of words are unaccented in the native and SJ vocabulary (Sibata 1994). According to the statistical work by Kubozono (1996) (see also Kubozono 2006a,b and Kubozono and Ohta 1998 for details), unaccentedness in loanwords has to do with the total phonological length of the word and its syllable structure. More specifically, unaccentedness is characteristic of four-mora words that end in a sequence of light syllables. Some examples are given in (48).

(48) a. LLLL
mo.na.ri.za 'Mona Lisa', a.ri.zo.na 'Arizona', yo.se.mi.te 'Yosemite'

b. HLL
ban.da.na 'bandana', ai.o.wa 'Iowa', dai.a.na 'Diana', kon.so.me 'consomme'

In statistical terms (Kubozono 2006a), about 50% of words with either of the two syllable structures in (48) are unaccented. More specifically, loanwords consisting of four light syllables as in (48a) show a stronger tendency towards the unaccented pattern than those that begin with a heavy syllable followed by light syllables: 54% vs. 45%. These ratios contrast very sharply with the low percentages of four-mora

unaccented words with other syllable structures: LHL (24%), LLH (19%) and HH (7%). Particularly striking is the difference in unaccented ratio between HLL (45%) and LLH (19%), which would have the same foot structure under foot-based analyses. These statistical results reinforce the argument that syllable structure plays a vital role in Japanese accentuation. This syllable-based analysis should be extended to cover loanwords of different lengths and also native and SJ words. This extended analysis will hopefully enable us to uncover the linguistic conditions under which unaccented words emerge in the Japanese lexicon.

Returning to the discussion of unaccented loanwords, a more accurate prediction can be made about the unaccentedness of loanwords if we allow for the distinction between epenthetic and non-epenthetic (or underlying) vowels. An orthodox idea about the relationship between vowel epenthesis and accent is that accent rules apply after epenthetic vowels have been inserted in loanwords. This idea can be substantiated by a number of words that bear an accent on an epenthetic vowel. Consider, for example, some of the words in (38), repeated below: < > means an epenthetic vowel.

(49) p<u>'.ra.s<u> 'plus'
k<u>.ri.s<u>'.ma.s<u> 'Christmas'

The words in (49) clearly indicate that epenthetic vowels are already present in the input to the antepenultimate rule in (37) or to any equivalent accent rule in Japanese. However, Japanese displays several phenomena in which epenthetic vowels are 'invisible', i.e., they behave as if they were not present. One such phenomenon concerns unaccented loanwords.

Note that all the words in (48) have a non-epenthetic vowel word-finally. In contrast, loanwords ending in an epenthetic vowel mostly follow the accent rule in (37) and attract an accent accordingly.

(50) a. LLLL s<u>.t<o>'.re.s<u> 'stress', p<u>.ro'.se.s<u> 'process', ba.ri'.u.m<u> 'Barium'
b. HLL a'n.de.s<u> 'Andes', ka'p.p<u>.r<u> 'couple', si'n.ba.r<u> 'cymbals'

If we exclude loanwords ending in an epenthetic vowel, the percentage of unaccented words with the two syllable structures in (48) reaches 90% – that is, most four-mora loanwords are unaccented if they end in a sequence of light syllables and do not involve an epenthetic vowel word-finally. This ratio is remarkably high, and actually higher than the average rate (about 60%) of unaccented four-mora words in the native and SJ strata.

The 'invisibility' of epenthetic vowels may not be so surprising since they do often show this kind of irregular behavior in other languages, too (see Alderete

1995; Michelson 1981, 1988; Potter 1994). However, the accent patterns we see in Japanese are particularly mysterious in several respects (Kubozono 2001b). In the first place, it is unclear why word-final epenthetic vowels are visible to the antepenultimate accent rule in the words in (49), while those in (50) are invisible to the rule responsible for unaccentedness.

Secondly, it remains a mystery why word-final epenthetic vowels in (50) contribute to the mora count of the antepenultimate rule in (37), while the same vowels are invisible to the rule responsible for unaccentedness. These questions remain for future work.

Loanword accentuation raises some more questions regarding the interaction between accent and epenthetic vowels (see Kubozono 2001b for more facts and mysteries). These questions, if properly addressed, will provide important insights into the nature of accent and its relationship with segments and syllable structure. This can potentially have a profound impact on analysis of the phonological structure of Japanese as well as on phonological theory in general, especially output-oriented Optimality Theory.

7.4 Accent of alphabetic acronyms

As the last case of loanword accent, let us consider the accentuation of alphabetic acronyms (or initialisms) including *PC* and *PTA*, which exhibit accent patterns remarkably different from ordinary loanwords in Tokyo Japanese. While many of these expressions come from English and other languages, quite a few are coined in Japanese itself such as *JR* (Japan Railways) and *NHK* (Nihon Hōsō Kyōkai, the national broadcasting corporation). These alphabetic acronyms are written in the English alphabet in Japanese books and newspapers and, moreover, their origin does not affect their accent patterns.

7.4.1 Accented acronyms

There are two accentual analyses in traditional descriptions, both reported in the appendix to *The Sanseido Shinmeikai Accent Dictionary* (1981 and 2001). Its 1981 version claims that alphabetic acronyms follow the rule in (51), while its 2001 version proposes the generalization in (52).

(51) The first mora of the last letter is accented.

(52) The most basic pattern is an unaccented pattern (i.e., flat pitch), although words ending in a long vowel or diphthong are accented on the penultimate mora.

Of these two generalizations, the second one falls into several difficulties (Kubozono 2003b). First, one finds many alphabetic acronyms that are accented on the penultimate mora regardless of their syllable structure. In fact, most three-letter and four-letter acronyms such as *NHK*, *BGM* and *YMCA* are accented on the last but one mora even though they do not end in a long vowel/diphthong. This is shown in (53).

(53) enu-eiti-ke'e (NHK), bii-zii-e'mu (BGM), esu-oo-e'su (SOS), wai-emu-sii-e'e (YMCA)

Second, and more important, acronyms that end in a three-mora or four-mora element are accented on the initial mora of this element rather than on its penultimate mora, as exemplified in (54). This productive pattern cannot be accounted for by the generalization in (52).

(54) zyee-a'aru (JR), dii-e'iti (DH), bui-e'kkusu (VX), bui-tii-a'aru (VTR), bii-emu-da'buryuu (BMW)

In comparison, the rule in (51) is capable of accounting for the patterns in both (53) and (54). However, this rule fails to cover unaccented acronyms such as those in (55), which seem to represent yet another general accent pattern of alphabetic acronyms.

(55) esu-eru (SL), oo-esu (OS), ehu-emu (FM)

Putting aside the unaccented pattern in (55) for a moment, the rule in (51) is capable of accounting for the accentuation of accented acronyms. The question is where this regularity comes from. Of the two accent patterns in (53) and (54), the pattern in (53) is substantially different from that of ordinary loanwords, which are mostly accented on the third or fourth mora from the end of the word if they are accented at all. Thus, the loanword /kja'n.dii/ 'candy' is accented on its initial syllable in Japanese, whereas the acronym /sii.di'i/ 'CD' and /e.mu.di'i/ 'MD' are accented on their final syllable. Similarly, the acronyms in (53) would receive the following accent patterns if they were subject to the loanword accent rule in (37).[20]

(56) e.nu.ei.ti'.kee (NHK)
bii.bi'i.sii (BBC)
e.su.o'o.e.su (SOS)
wai.e.mu.si'i.ee (YMCA)

20 *NHK* and *SOS* exceptionally permit these accent patterns alongside the patterns in (53), but other acronyms do not exhibit such variation.

One way of accounting for the accent patterns in (53) is to assume that Japanese acronyms have borrowed or copied the original stress pattern of English, where acronyms usually receive primary stress on the initial syllable of their final member: e.g., PĆ, PTÁ, BBĆ. While this account seems simple, it cannot explain the difference between alphabetic acronyms and ordinary loanwords, that is, why the former follows the accent pattern of the source words, while the latter do not: e.g., /ba'.na.na/ 'banana' and /wa.si'n.ton/ 'Washington'. Moreover, the English-based account cannot provide a principled account of the cross-dialectal differences in the accentuation of alphabetic acronyms, as we will see shortly below.

A more plausible account for the accent patterns in (53) is to attribute them to the compound accent rule of the language (Kubozono 2010). As mentioned in section 7.2 above, compound words in Tokyo Japanese tend to preserve the lexical accent of their final member as the compound accent. Since alphabetic letters are all pronounced with an initial accent when pronounced in isolation in Tokyo Japanese, the compound accent rule of this dialect predicts that the initial syllable of the final member will bear accent in a compound. This prediction is fully borne out as we saw in (53) and (54) above. (57) illustrates this point by comparing the accent patterns of acronyms (57a) with those of compound nouns whose final member is an ordinary loanword (57b). Since (57a) and (57b) exhibit parallel accentual behaviors, (57a) can be accounted for by the same accent rule that is responsible for (57b).

(57) a. bi'i + bi'i + si'i → bii-bii-si'i 'BBC'
pi'i + si'i → pii-si'i 'PC'
zye'e + a'a.ru → zyee-a'a.ru 'JR'

b. di'.zu.nii + si'i → di.zu.nii-si'i 'Disney Sea'
u.ru.to.ra + si'i → u.ru.to.ra-si'i 'Ultra C (very difficult performance)'
zyu'u + a'a.ru → zyuu-a'a.ru '10 ares (area)'

This alternative analysis has a further advantage of accounting for the accent patterns of alphabetic acronyms in other dialects (Kubozono 2010). Kagoshima Japanese, for example, exhibits remarkably different accent patterns from Tokyo Japanese for alphabetic acronyms. In general, this dialect permits only two accent patterns – called Type A and Type B (Hirayama 1951) – which are differentiated from each other with respect to the position of high pitch: Type A bears high pitch on the penultimate syllable, whereas Type B has high pitch on the final syllable. This dialect has a compound accent rule that is a mirror-image of that of Tokyo Japanese in that the accent pattern (Type A or B) of the initial member spreads over the entire compound. This is exemplified in (58), where capital letters denote high-pitched portions. Thus, /na.tu-ja.SU.mi/ 'summer holiday' is high-pitched on the penultimate syllable since its initial member, /NA.tu/, is a Type A morpheme. Likewise, /ha.ru-ja.su.MI/ takes the Type B pattern since its initial member, /ha.RU/ is a Type B morpheme.

(58) a. NA.tu + ya.su.MI → na.tu-ya.SU.mi 'summer, holiday: summer holiday'
b. ha.RU + ya.su.MI → ha.ru-ya.su.MI 'spring, holiday; spring holiday'

Alphabetic acronyms also fall into two accent groups. Some show the Type A pattern, as in (59a), whereas others take the Type B pattern, as in (59b).

(59) a. e.hu-E.mu 'FM'
e.RU-pii 'LP'
e.hu-BII-ai 'FBI'

b. ee-e.MU 'AM'
ee-PII 'AP'
sii-ai-EE 'CIA'

A careful examination of the data reveals that the two accent patterns in (59) reflect the accent differences of the initial members. Thus, the acronyms in (59a) take the Type A pattern since their initial members, i.e., 'F' and 'L', are pronounced with the Type A pattern – i.e., high pitch on the penultimate syllable – in citation form. Likewise, those in (59b) show the Type B pattern – high pitch on the final syllable – since their initial members, i.e., 'A' and 'C', are Type B morphemes. This compound accent effect is illustrated in (60).

(60) a. E.hu + E.mu → e.hu-E.mu
E.ru + PII → e.RU-pii
E.hu + BII + AI → e.hu-BII-ai

b. EE + E.mu → ee-e.MU
EE + PII → ee-PII
SII + AI + EE → sii-ai-EE

Kagoshima Japanese is crucially different from Tokyo Japanese in two respects: the accentuation of alphabetic letters and the content of the compound accent rule. Yet, both dialects show a crucial similarity in obeying their own compound accent rules in the accentuation of alphabetic acronyms. This generalization holds in other dialects, too (Kubozono 2010). In sum, accented alphabetic acronyms follow the compound accent rule across dialects.

7.4.2 Unaccented acronyms

In Tokyo Japanese, some alphabetic acronyms show the unaccented pattern as mentioned in (55) above. Their distribution can be predicted by and large on the basis of their phonological structure, just as unaccented loanwords occur in highly

predictable contexts (see section 7.3 above). Interestingly, the unaccented pattern emerges in alphabetic acronyms in the same contexts where unaccented loanwords occur. First, the unaccented pattern is observed only in four-mora acronyms. Since alphabetic letters are at least two moras long in citation form, this means that the unaccented pattern occurs only in two-letter acronyms in which each member consists of two moras: e.g., /bii-e.su/ 'BS', /sii-e.mu/ 'CM', /e.su-e.ru/ 'SL'. Seen conversely, the accent pattern in question does not occur in five-mora or longer acronyms: e.g., /zjee-a'a.ru/ 'JR', /pii-a'a.ru/ 'PR', /dii-e'i.ti/ 'DH', /ek.ku.su-pi'i/ 'XP', /bii-zii-e'.mu/ 'BGM'. This explains why /bii-e.su/ 'BS', /oo-e.su/ 'OS' and /zii-e.mu/ 'GM' are unaccented while /tii-bii-e'.su/ 'TBS', /e.su-oo.e'.su/ 'SOS' and /bii-zii-e'.mu/ 'BGM' are not. In Kubozono's (2003b) statistical data, there is no instance of five-mora or longer acronyms that takes the unaccented pattern. In contrast, many four-mora acronyms are unaccented.

A second factor responsible for the unaccented pattern in alphabetic acronyms concerns their prosodic structure. Just like ordinary loanwords such as /mo.na.ri.za/ 'Mona Lisa' and /kon.so.me/ 'consomme', the unaccented pattern is observed predominantly in four-mora acronyms that end in a sequence of light syllables. In fact, the unaccented pattern is found only in four-mora acronyms whose final member consists of two light syllables: 80% of acronyms with this prosodic structure take the accent pattern in question (Kubozono 2003b). In comparison, no four-mora acronyms that end in a heavy syllable take the unaccented pattern. This explains why /bii-e.su/ 'BS', /ee-e.mu/ 'AM' and /zii-e.mu/ 'GM' are unaccented, while /e.su-bi'i/ 'SB', /e.mu-e'e/ 'MA' and /e.nu-zi'i/ 'NG' are accented.

In addition, four-mora acronyms ending in a sequence of light syllables become unaccented to different degrees depending on the prosodic structure of their initial member. If this member consists of two light syllables, virtually all acronyms are unaccented. In contrast, the ratio of the unaccented pattern goes down to 70% if the initial member consists of one heavy syllable. This is responsible for the fact that four-mora acronyms consisting of four light syllables, e.g., /e.hu-e.mu/ 'FM', /e.su-e.ru/ 'SL', /e.ru-e.ru/ 'LL', are almost invariably unaccented, while those consisting of a heavy syllable followed by two light syllables often show variation between the accented and unaccented patterns: e.g., /ee-e'.mu/~/ee-e.mu/ 'AM', /oo-e'.ru/~/oo-e.ru/ 'OL', /bii-e'.su/~/bii-e.su/ 'BS'. This additional factor is also shared by ordinary loanwords as mentioned in section 7.3 above.

Incidentally, the nature of the word-final vowel does not seem relevant in the accentuation of alphabetic acronyms. In fact, all acronyms with a final light syllable end in an epenthetic vowel, i.e., a vowel inserted during the process of loanword adaptation: e.g., /e.h<u>-e.m<u>/ 'FM', /e.s<u>-e.r<u>/ 'SL', /e.r<u>-e.r<u>/ 'LL', /ee-e.m<u>/ 'AM', /oo-e.r<u>/ 'OL'. Nevertheless, these acronyms readily become unaccented. This contrasts with the fact mentioned in section 7.3 above, that ordinary loanwords are resistant to the unaccented pattern if they end in an epenthetic vowel. Apart from this minor difference, however, alphabetic acronyms take the unaccented

pattern in basically the same phonological contexts as ordinary loanwords. The accent pattern in question is highly predictable in both types of words.

8 Conclusion

In this chapter we saw many phonological phenomena involving loanwords as well as the processes that are responsible for them. In section 2, we examined how Japanese borrows vowels and consonants from English and other foreign languages. In section 3, we analyzed the strategies that Japanese employs to choose epenthetic vowels. Section 4 examined the solutions the language relies on to avoid creating hiatus, or the marked structure involving vowel-vowel sequences across a syllable boundary. Section 5 looked at the asymmetries between /ai/ and /au/, which exhibit contrastive but consistent behaviors in several independent processes. Section 6 examined various phenomena that conspire to avoid superheavy, i.e., trimoraic, syllables. Section 7 discussed loanword accent with main focus on the accent patterns of simplex loanwords and alphabetic acronyms.

By looking at these various aspects of loanword phonology, we have seen the importance of studying loanwords for a better understanding of the host language. For one thing, loanword phonology serves as a mirror that reflects the structure of the host language that would not otherwise show itself clearly. The discussion in section 6, for example, showed that a constraint prohibiting trimoraic syllables is at work in Japanese, too, although it cannot be seen clearly in native and SJ words. Moreover, loanword processes often reveal the basic nature of Japanese phonology. For example, what has been formulated as the loanword accent rule is, in fact, a rule of accented native words that can be generalized with compound accent rules on the one hand, and with the Latin accent rule on the other (section 7).

The present study has uncovered not only many basic phonological structures of Japanese but also many new questions for future work. We saw some of them in the preceding sections, but there are more questions that remain unsolved. For example, the analysis of alphabetic acronyms in section 7 reveals a mystery about the nature of this type of word. The discussion there showed that the two accent types of alphabetic acronyms in Tokyo Japanese – accented and unaccented – are highly predictable from their phonological structures. It also revealed that acronyms are subject to the compound accent rule when they are accented. Given this analysis, a question naturally occurs as to why four-mora acronyms ending in a sequence of light syllables fail to undergo the compound accent rule or, to be more precise, why they are not analyzed as compounds phonologically. For example, /e.su-bi'i/ 'SB' and /tii-bii-e'.su/ 'TBS' are processed as compound nouns and thereby retain the lexical accent of the final member due to the compound accent rule (see (57) above). On the other hand, /bii-e.su/ 'BS' does not undergo this phonological rule but is processed

as if it were a non-compound word just like ordinary loanwords that are unaccented, e.g., /kon.so.me/ 'konsomme'. It seems difficult to reconcile the compound accent analysis proposed for accented acronyms with the unaccented pattern shown by four-mora acronyms with a particular prosodic structure. The present study has revealed this and many other questions of a similar kind that remain for future work.

Acknowledgments

I would like to thank Donna Erickson, Mark Irwin, Itsue Kawagoe, and Tim Vance for their invaluable comments on the draft of this chapter. All errors that remain are my own. The work reported in this chapter was supported by the NINJAL collaborative research project 'Phonological characteristics of the Japanese lexicon' and the JSPS grants-in-Aid for Scientific Research (Grant no. 22242011 and 25580098).

References

Akinaga, Kazue. 1981. Akusento shūtoku hōsoku [Laws of accent learning]. In the appendix to *Meikai nihongo akusento jiten* [Japanese accent dictionary]. Tokyo: Sanseidō.

Alderete, J. 1995. Faithfulness to prosodic heads. Unpublished ms., University of Massachusetts, Amherst.

Árnason, Kristjan. 1980. *Quantity in historical linguistics*. Cambridge: Cambridge University Press.

Beckman, Jill. 1998. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*. Amherst, MA: University of Massachusetts dissertation. (Garland Press, 1999).

Casali, Roderic F. 1996. *Resolving hiatus*. Los Angeles: University of California dissertation. (Garland Press, 1998).

Casali, Roderic F. 2011. Hiatus resolution. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1434–1460. Malden, MA and Oxford: Wiley-Blackwell.

Choi, Kyung-Ae. 2000. An overview of Korean loanword phonology. *Journal of the Phonetic Society of Japan* 6(1). 22–33.

Crawford, Cliffod. 2007. The role of loanword diffusion in changing adaptation patterns: A study of cornonal stops in Japanese borrowings. *Working Papers of the Cornell Phonetics Laboratory* 16. 32–56.

Dohlus, Katrin. 2005. On the asymmetry in the adaptation pattern of German umlaut in Japanese. *Journal of the Phonetic Society of Japan* 9(1). 39–49.

Dohlus, Katrin. 2010. *The role of phonology and phonetics in loanword adaptation: German and French front rounded vowels in Japanese.* (Europaische Hochschulschriften. Reihe 21: Linguistik). Frankfurt am Main: Peter Lang.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono, ed., *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Hall, Nancy. 2011. Vowel epenthesis. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1576–1596. Malden, MA and Oxford: Wiley-Blackwell.

Hayasi, T. 1982. Nihongo no kanjion [Sino-Japanese pronunciations in Japanese]. In Norio Nakata (ed.), *Nihongo no sekai 4: Nihon no kanji* [The world of Japanese 4: Chinese characters in Japan]. Tokyo: Chūōkōronsha.

Hayes, Bruce. 1986. Inalterability in CV Phonology. *Language* 62(2). 321–351.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20(2). 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hirayama, Teruo. 1951. *Kyūshū hōgen onchō no kenkyū* [Studies on the tone of Kyushu dialects]. Tokyo: Gakkai no shishinsha.

Irwin, Mark. 2011. *Loanwords in Japanese*. Amsterdam/Philadelphia: John Benjamins.

Ito, Emi. 2000. Gairaigo no fukugōgo ni okeru tokushu haku no arawarekata [Behavior of special moras in the truncation of loanword compounds]. *Nihon Joshidaigaku Kiyō* 49. 19–29. Nihon Women's University.

Ito, Junko. 1986. *Syllable theory in Prosodic Phonology.* Amherst, MA: University of Massachusetts dissertation. (Garland Press, 1988).

Ito, Junko. 1990. Prosodic minimality in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.

Ito, Junko, Yoshihisa Kitagawa and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Ito, Junko and Armin Mester. 1995. Binarity. Paper presented at the GLOW phonology workshop, Tromso, Norway, 3 June.

Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takeshi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic Studies*, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono, ed., *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Jakobson, Roman. 1968. *Child Language, aphasia and phonological universals.* The Hague: Mouton.

Jones, Daniel. 1960. *An outline of English phonetics*. Cambridge: Heffer and Sons.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology.* Santa Cruz, California: University of California dissertation.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82(3). 536–574.

Kawahara, Shigeto. 2011. Aspects of Japanese loanword devoicing. *Journal of East Asian Linguistics* 20(2). 169–194.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono, ed., *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to the sounds of Japanese]. Tokyo: Ōfūsha.

Kindaichi, Kyosuke. 1976. *Nihongo no hensen* [Historical development of Japanese]. Tokyo: Kodansha.

Kubozono, Haruo. 1995a. *Gokeisei to on'in kōzō* [Word formation and phonological structure]. Tokyo: Kurosio.

Kubozono, Haruo. 1995b. Constraint interaction in Japanese phonology: Evidence from compound accent. In Rachel Walker, Ove Lorentz and Haruo Kubozono (eds.), *Phonology at Santa Cruz, Volume 4: Papers on Stress, Accent, and Alignment*, 21–38.

Kubozono, Haruo. 1996. Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin* (The Phonetic Society of Japan) 211. 71–82.
Kubozono, Haruo. 1997. Lexical markedness and variation: A non-derivational account. *WCCFL* 15, 273–287.
Kubozono, Haruo. 1999a. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.
Kubozono, Haruo. 1999b. *Nihongo no onsei* [The sounds of Japanese]. Tokyo: Iwanami.
Kubozono, Haruo. 2000. Nihongo no goi to onsetsu kōzō [Japanese vocabulary and syllable structure]. *Nihongo Kenkyū* [Studies on Japanese] 20. 1–18. Tokyo Metropolitan University.
Kubozono, Haruo. 2001a. On the markedness of diphthongs. *Kobe Papers in Linguistics* 3. 60–73.
Kubozono, Haruo. 2001b. Epenthetic vowels and accent in Japanese: Facts and paradoxes. In Jeroen van de Weijer and Tetsuo Nishihara (eds.), *Issues in Japanese phonology and morphology*, 113–142. Berlin and New York: Mouton de Gruyter.
Kubozono, Haruo. 2002. Temporal neutralization in Japanese. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 171–201. Berlin: Mouton de Gruyter.
Kubozono, Haruo. 2003a. The syllable as a unit of prosodic organization in Japanese. In Caroline Fery and Ruben van der Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.
Kubozono, Haruo. 2003b. Accent of alphabetic acronyms in Tokyo Japanese. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.), *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*, 356–370. Tokyo: Kaitakusha.
Kubozono, Haruo. 2004a. What does Kagoshima Japanese tell us about Japanese syllables? In Taro Kageyama and Hideki Kishimoto (eds.), *Nihongo no bunseki to gengo ruikei* [Analysis of Japanese and language typology], 75–92. Tokyo: Kurosio.
Kubozono, Haruo. 2004b. On'in kōzō kara mita tanjungo to gōseigo no kyōkai [Phonological boundary between simplex and compound words]. In Spoken Language Working Group (ed.), *Speech and grammar* 4, 123–143. Tokyo: Kurosio.
Kubozono, Haruo. 2005. [ai]-[au] asymmetry in English and Japanese. *English Linguistics* 22(1). 1–22.
Kubozono, Haruo. 2006a. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.
Kubozono, Haruo. 2006b. *Akusento no hōsoku* [*Laws of accent*]. Tokyo: Iwanami.
Kubozono, Haruo. 2007a. Tonal change in language contact: Evidence from Kagoshima Japanese. In Tomas Riad and Carlos Gussenhoven (eds.), *Tones and tunes Volume 1: Typological studies in word and sentence prosody*, 323–351. Berlin: Mouton de Gruyter.
Kubozono, Haruo (ed.). 2007b. Special issue on *sokuon*, or moraic obstruent. *Journal of the Phonetic Society of Japan* 11(1).
Kubozono, Haruo. 2008a. [ai]-[au] asymmetry: A phonetic account. In Haruo Kubozono (ed.), *Asymmetries in phonology: An East-Asian perspective*, 147–163. Tokyo: Kurosio.
Kubozono, Haruo. 2008b. Japanese accent. In Shigeru Miyagawa and Mamoru Saito (eds.), *Handbook of Japanese linguistics*, 165–191. Oxford: Oxford University Press.
Kubozono, Haruo. 2010. Accentuation of alphabetic acronyms in varieties of Japanese. *Lingua* 120 (10). 2323–2335.
Kubozono, Haruo. 2011. Japanese pitch accent. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 5*. 2879–2907. Malden, MA and Oxford: Wiley-Blackwell.
Kubozono, Haruo. 2013a. Japanese word accent. In Mark Aronoff (ed.), *Oxford Bibliographies in Linguistics* (online). New York: Oxford University Press.
Kubozono, Haruo (ed.). 2013b. Special issue on Japanese geminate obstruents. *Journal of East Asian Linguistics* 22(4).

Kubozono, Haruo. this volume. Introduction. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo, Junko Ito and Armin Mester. 1997. On'inkōzō kara mita go to ku no kyōkai: Fukugō meishi akusento no bunseki [The phonological boundary between the word and the phrase: An analysis of Japanese compound noun accentuation]. In Spoken Language Working Group (ed.), *Bunpō to onsei* [*Speech and grammar*], 147–166. Tokyo: Kurosio.

Kubozono, Haruo and Satoshi Ohta. 1998. On'in kōzō to akusento [Phonological structure and accent]. Tokyo: Kenkyusha.

Kubozono, Haruo and Takeru Honma. 2002. *Onsetsu to mōra* [The syllable and the mora]. Tokyo: Kenkyusha.

Kuwamoto, Yuji. 1998a. Nihongo no ryakugo no on'in kōzō [The phonological structure of clipped words in Japanese]. *On'in Kenkyū* [Phonological Studies] 1. 161–168.

Kuwamoto, Yuji. 1998b. Ryakugo kara mita nihongo no chōjū onsetsu no kōzō ni tsuite [Truncation and the structure of trimoraic syllables in Japanese]. *Gengo Kagaku Ronshū* 2. 25–35. Tohoku University.

Lovins, Julie. 1975. *Loanwords and the phonological structure of Japanese*. Indiana University Linguistics Club.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Michelson, Karin. 1981. Stress, epenthesis and syllable structure in Mohawk. In George N. Clements (ed.), *Harvard studies in phonology* 2, 311–353. Indiana University Linguistics Club.

Michelson, Karin. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer.

Moore, Samuel and Albert H. Marckwardt. 1981. *Historical outlines of English sounds and inflections*. Ann Arbor: George Wahr Publishing Company.

Mutsukawa, Masahiko. 2009. *Japanese loanword phonology: The nature of inputs and the loanword sublexicon*. Tokyo: Hituzi Syobo.

Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Nishimura, Kohei. 2003. Lyman's Law in loanwords. Nagoya University MA Thesis.

Paradis, Carole and Jean-François Prunet. 2000. Nasal vowels as two segments: Evidence from borrowings. *Language* 76(2). 324–357.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Poser, William. 1988. Glide formation and compensatory lengthening in Japanese. *Linguistic Inquiry* 19. 494–502.

Potter, B. 1994. Serial optimality in Mohawk prosody. *Proceedings of the Annual Meeting of the Chicago Linguistic Society* 30, 347–61.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Technical Report #2, Rutgers Center for Cognitive Science, Rutgers University. (Blackwell, 2004).

Quackenbush, Hiroko and Mieko Ohso. 1990. *Gairaigo no keisei to sono kyōiku* [Word formation and teaching of loanwords]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

Sagisaka, Yoshinori and Yoh'ichi Tohkura. 1984. Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyo kisoku [Phoneme duration control for speech synthesis by rule]. *Denshi Tsūshin Gakkai Ronbunshi* [The Transactions of the Institute of Electronics, Information and Communication Engineers A] 67(7). 629–636.

Sato, Hirokazu. 2002. Gairaigo ni okeru onsetsu fukugō e no kubunka to akusento [Decomposition into syllable complexes and the accenting of Japanese borrowed words]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 6(1). 67–78.
Sherer, Tim D. 1994. *Prosodic phonotactics*. Amherst, MA: University of Massachusetts dissertation.
Sibata, Takesi. 1994. Gairaigo ni okeru akusento kaku no ichi [The location of accent in loanwords]. In Kiyoji Sato (ed.), *Kokugo ronkyū 4: Gendaigo hōgen no kenkyū* [Studies in Japanese 4: Studies in modern Japanese and dialects], 338–418. Tokyo: Meiji Shoin.
Shinohara, Shigeko. 2000. Default accentuation and foot structure in Japanese: Evidence from Japanese adaptations of French words. *Journal of East Asian Linguistics* 9. 55–96.
Stevens, Ken. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3–45.
Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13. 209–235.
Uffmann, Christian. 2006. Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua* 116. 1079–1111.
Vance, Timothy J. and Yuka Matsugu. 2008. The /y/-/w/ asymmetry in Japanese loanwords. In Haruo Kubozono (ed.), *Asymmetries in phonology: An East-Asian perspective*, 139–146. Tokyo: Kurosio.
Yavaş, Mehmet. 1998. *Phonology: Development and disorders*. San Diego: Singular Publishing Group.
Zec, Draga. 1995. The role of moraic structure in the distribution of segments within the syllable. In Jacques Durand and Francis Katamba (eds.), *Frontiers of phonology: Atoms, structures, derivations*, 149–179. New York: Longman.

Junko Ito and Armin Mester

9 Word formation and phonological processes

1 Introduction

The goal of this chapter is to outline the major types of word formation in Japanese from a phonological perspective.[1] In addition to laying out the main rules and generalizations, with an emphasis on phonological patterns and prosodic constraints, we will summarize previous work and sketch new developments. Theories focusing on the phonological aspects of word formation – in particular, Lexical Phonology (Kiparsky 1982) and Prosodic Morphology (McCarthy and Prince 1990b, as well as more recent optimality-theoretic developments of these theories, such as Stratal OT (Kiparsky 2000) and Optimal Interleaving (Wolf 2008)) – have led to a significantly deeper understanding of word formation and morphological structure. While covering various phonological details of Japanese word formation – affixation in sections 2 and 3, compounding in section 4, and templatic morphology in section 5 –, it is our aim to focus on general theoretical ramifications of the point under discussion, and highlight both how general phonological theory has informed the analysis of Japanese word formation in the past, and how phonological studies of different types of Japanese word formation have contributed important case studies leading to advances in the general theory of phonology, and phonology-morphology interactions.

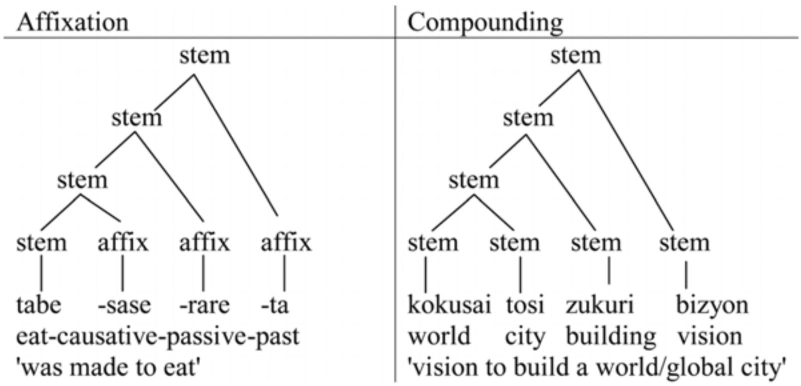
2 Phonology of affixation

In affixation structures, a stem and an affix together form a larger stem, to which another affix can be attached to form another even larger stem, as long as semantic and selectional restrictions are obeyed. The diagrams in (1) illustrate how affixation is structurally parallel to compounding (discussed in section 4), the difference being that each complex stem is composed of *stem+affix* in the former, *stem+stem* in the latter.[2]

1 For the morphosyntactic aspects of word formation, see the *Handbook of Japanese Lexicon and Word Formation* edited by Taro Kageyama in this series.

2 In order to avoid unnecessary terminological clutter, we have opted for the simple morphological bifurcation between *stem* (any morphological complex based on a root) and *affix* (bound form). We do not distinguish here between derivational and inflectional affixes, since they do not exhibit distinct phonological properties in Japanese. Unless a distinction is called for, the neutral term *stem* refers to both roots and affixed forms, as shown in (1).

(1)



Standard Japanese language dictionaries (including those for foreign language learners, such as Kenkyusha's *Japanese-English Learner's Dictionary*) usually provide appendix charts of verbal and adjectival derivational and inflectional paradigms (listing the root, present, formal present, negative, inchoative, gerundive, past, etc.), and accent dictionaries (e.g., NHK's *Nihongo Hatsuon Akusento Jiten* or Sanseido's *Shinmeikai Akusento Jiten*) devote several pages to the varying accentual patterns associated with different derivational and inflectional suffixes (see Kawahara, Ch. 11, this volume). The morphophonemics of the paradigms of inflected words (in Japanese, only verbs and adjectives have such paradigms, not nouns) have been studied in different frameworks from the earliest structural and generative traditions (Bloch 1946a,b; Martin 1952; McCawley 1968; Hattori 1973; de Chene 2010; Davis and Tsujimura 1991; Ito and Mester 2004; Sano 2012). Rather than attempting to summarize these works, we present the core phonological patterns observed in these paradigms, and point out where and how they bear on phonological theories and universals.

2.1 Preliminaries: phonological typology of stems and suffixes

Verbal stems come in two phonological varieties, those ending in a consonant versus those ending in a vowel (2).[3]

(2)

C-final stems:				V-final stems:			
nom-	'drink'	kosur-	'rub'	tabe-	'eat'	nobi-	'stretch'
kik-	'hear'	tat-	'stand'	mi-	'see'	kurabe-	'compare'
moraw-	'receive'	oyog-	'swim'	tome-	'stop'		
tob-	'fly'	hatarak-	'work'				

3 In Japanese school grammar terminology, the C-final stems correspond to verbs with *godan katsuyō* '5-vowel conjugation', and the V-final stems correspond to verbs with *kami-ichidan katsuyō* 'i-conjugation' and *shimo-ichidan katsuyō* 'e-conjugation' (see below for some discussion).

Dictionaries list verbs in the present indicative form with the *-ru/-u* ending, and V-final stems are often referred to as *ru*-verbs (e.g., tabe*ru*, mi*ru*) and C-final stems as *u*-verbs (e.g., nom*u*, kik*u*).[4] They are easily identifiable as listed in the dictionary, with one caveat. Not all forms ending in the sequence /ru/ are ru-verbs, because the /r/ can also be the final consonant of the stem. Thus the stem for 'understand' in *wakaru* is /wakar/, not */waka/. Synchronically, all V-final verbal stems end in a front vowel (/i/ or /e/),[5] so a verb with any other vowel in its last syllable must be C-final (e.g., *suw**ar**-u* 'sit-present', *mam**or**-u* 'protect-present', *kos**ur**-*u 'scrub-present'). This is only a one-way implication (back vowel /u,o,a/ → C-stem): Front vowels with /r/ occur both in V-final stems or C-final stems, leading to (segmental) homonyms in the present tense as in (3) (sometimes differing in accent, see section 3.4 below), with different morphological junctures. In other parts of the paradigm (such as the negative present show below), the two stems show different formations.

(3)

	PRESENT	gloss	cf. NEGATIVE PRESENT
a.	ki-ru	'wear'	ki-nai
	ki'r-u	'cut'	kir-a'nai
b.	kae-ru	'change'	kae-nai
	ka'er-u	'return'	kaer-a'nai
c.	i-ru	'be, exist'	i-nai
	ir-u	'need'	ir-anai

Using a list available on the internet of the most common Japanese verbs[6] (containing approximately 500 items), we find 312 (63%) ending in the sequence /ru/. 145 (46%) of these are V-stems, and 167 (54%) *r*-final C-stems.

(4) Verbs ending in the sequence /ru/

r-final (…*Vr-u*)			V-final (…*V-ru*)		
ar	53	32%	e	128	88%
ir	51	30%	i	17	12%
er	30	18%	u	–	
or	20	12%	o	–	
ur	13	8%	a	–	
Total	167		Total	145	

4 See section 2.2 below for the *-ru/-u* allomorphy.

5 This restriction already goes back to Old Japanese.

6 From http://wiki.verbix.com/Verbs/JapaneseVerbList with 492 verbs, checked against http://www.japaneseverbconjugator.com/JVerbList.asp with 418 verbs. (Verbix contains verb conjugations for many regional, national and international languages, including Japanese. Japaneseverbconjugator hosts a Japanese verb database coded to conjugate verbs in different tenses.)

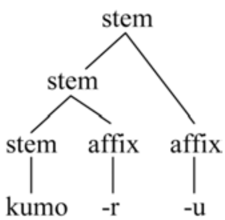
Overall, 348/493 verbs (71%) are C-final, and among the C-final verbs, the vast majority (48%) are *r*-final (5), compared to the closest contender, *s*-final verbs (14.9%).

(5) C-final verbs (total: 348)

r-final	167	48.0%	b-final	12	3.4%
s-final	52	14.9%	g-final	10	2.9%
w-final	36	10.3%	t-final	7	2.0%
k-final	32	9.2%	n-final	1	0.3%
m-final	31	8.9%			

This preponderance of *r*-final verb stems is due to the existence of *r*-final derivational suffixes, such as the deadjectival /-mar/ (e.g., *kata-mar-u* 'hard-en', see Martin 1952: 60–61 for a list and Bloch 1946b for details), as well as the mono-consonantal stem-forming /-r/ (6).

(6) Stem-forming r-suffix



kumo	'cloud'	kumo-r-u	'to become clouded'
kage	'shadow'	kage-r-u	'to be shaded'
nezi	'screw'	nezi-r-u	'to screw'
guti	'grumble'	guti-r-u	'to grumble'
yazi	'jeer'	yazi-r-u	'to jeer at'
dozi	'mess'	dozi-r-u	'to mess up'
hosoi	'thin'	hoso-r-u	'to become thin'
hutoi	'fat'	huto-r-u	'to become fat'
biyooin	'beauty parlor'	biyo-r-u	'to go to the biyōin (beauty parlor)'
kokuhaku	'confession'	koku-r-u	'to confess (love)'
makudonarudo	'McDonald's'	maku-r-u	'to eat at McDonald's'

/-r/ is here part of the verb stem that the inflectional suffixes attach to, as shown by the fact that the conjugation paradigm keeps it intact (*kager-anai*, not **kagenai*,

nezir-anai, not **nezi-nai*, etc.). The stem-forming *r*-suffix is semi-productive: *X-r-u* formations are used alongside *X–suru* 'do X' compounds, where X is a loanword. Thus one finds *memo-r-u* alongside *memo-suru* 'to jot down a memo', lit. 'to do a memo.' We return to this verb-forming *r*-suffix below in section 3.4.

Verbal suffixes also come in two phonological varieties: mnemonically, C/V-suffixes and T-suffixes (7) (this classification goes back to Bloch (1946a), who calls the latter "stopped endings"). C/V-suffixes have both a C-initial allomorph and a V-initial allomorph, with the exception of the infinitive suffix, whose allomorphs are /-Ø, -i/. The T-suffixes invariably start with the consonant /t/ (or its voiced variant /d/, depending on the phonological environment), and they lack V-initial allomorphs.

(7)

C/V-suffixes:		T-suffixes:	
-sase, -ase	CAUSATIVE	-te, -de	GERUNDIVE
-rare, -are	PASSIVE	-ta, -da	PAST
-na, -ana	NEGATION	-tari, -dari	ALTERNATIVE
-ru, -u	PRESENT	-tara, -dara	CONDITIONAL ('if')
-yoo, -oo	VOLITIONAL	-tatte, -datte	CONCESSIVE CONDITIONAL ('even if')
-ro, -e	IMPERATIVE		
-reba, -eba	PROVISIONAL		
-rare, -e	POTENTIAL		
-Ø, -i	INFINITIVE		

C/V-suffixation and T-suffixation, in different ways, show allomorphy effects on the suffixes themselves, and/or the stems to which they attach. Most importantly, for the purposes of this chapter, these effects reveal the syllable structural constraints and segmental patterns of the Japanese phonological system.

2.2 The morphophonology of C/V-suffixes

The allomorphy of the C/V-suffixes is syllable-conditioned, that is, the choice of allomorph depends on the phonological shape of the stem to which they attach. A C-stem (e.g., *nom-* 'drink') is followed by a V-initial allomorph (*nom-ase*, **nom-sase*), whereas a V-stem (e.g., *tabe-* 'eat') is followed by a C-initial allomorph (*tabe-sase,* **tabe-ase*).

(8)

	V-stem (tabe-)	C-stem (nom-)
C-allomorph (-sase)	tabe-sase	*nom-sase
V-allomorph (-ase)	*tabe-ase	nom-ase

The allomorphy can be understood as being guided by the two most basic universal syllable structure constraints, ONSET (requiring onsets) and NOCODA (disallowing

codas).[7] Both allomorphs are available for all verbs, but the wrong choice of allomorph will lead to a violation of one of these constraints, e.g., *[.no*m*.sa.se.] violates NoCoda, and *[.ta.be.*a*.se.] violates Onset.[8]

Relying on Onset and NoCoda to choose the relevant allomorph raises the question regarding the status of these constraints in Japanese. The syllable structure of Old Japanese is considered to be [CV] (see Takayama, this volume), with obligatory onsets (except word-initially) and without codas, but modern Japanese has many words with vowel hiatus (e.g., *a.o* 'blue'), violating Onset, as well as words with certain kinds of codas (e.g., ki*t*.te 'stamp', to*n*.de 'fly-GERUND'), violating NoCoda. In optimality-theoretic terms, this means that Onset and NoCoda are high-ranking (and hence virtually unviolated) in Old Japanese, but low-ranking (with rampant surface violations) in Modern Japanese. Their low-ranking status, however, does not mean that they are not part of the synchronic grammar. Even if mostly inert, and not triggering any phonological alternations, they are still active in choosing between existing allomorphs (see Mascaró 1996 for a general theory of allomorphy in OT, and Ito and Mester 2004 for an OT analysis of Japanese verbal allomorphy).

Purely phonological analyses (deleting and/or inserting consonants and vowels) have been proposed in the generative tradition, with ordered segmental rules (Kuroda 1965; McCawley 1968) or with autosegmental spreading and delinking (Davis and Tsujimura 1991). The rules and conditions needed to derive the surface forms, however, are necessarily construction-specific, pertaining only to the verbal conjugations.[9] Thus outside of this morphological domain, we find not consonant deletion, as in putative /nom-sase/ → [nomØase], but rather vowel epenthesis, which would yield *[nom*u*sase]. Employing some version of level-ordered lexical phonology (Kiparsky 1982) may be a possible way out, but complications still arise since a level would have to be posited that is not only specific to the verbal paradigm but also to the particular verbal affixes that undergo certain changes and not others. The OT allomorph analysis has the advantage of relying on phonological constraints for the choice of allomorphs, without having to treat what has arisen historically as synchronic phonological processes (see de Chene 2010 for further discussion).

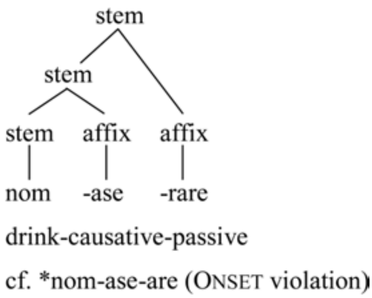
Affixed forms, such as [[*tabe*]*sase*] EAT-CAUSATIVE 'force to eat' or [[*nom*]*ase*] DRINK-CAUSATIVE 'force to drink', are themselves also stems, and as such allow further affixation [[[*tabe*]*sase*]*rare*] EAT-CAUSATIVE-PASSIVE 'be forced to eat', [[[*nom*]*ase*]*rare*] DRINK-CAUSATIVE-PASSIVE 'be forced to drink'. For the latter form, as depicted in (9a), even though the verbal root /nom/ is a C-stem, the affixed form [[*nom*]*ase*]

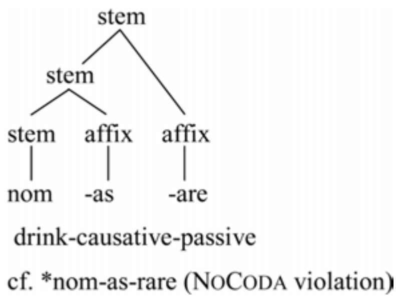
7 In terms of the elementary syllable theory of Prince and Smolensky (1993 [2004]).

8 The infinitive endings are /Ø, -i/ (7), where the null variant patterns with the C-allomorphs: *tabeØ(-ni iku)* '(go out) to eat' and *nom-i(-ni iku)* '(go out) to drink', and not **tabe-i* with an Onset violation, and **nomØ* with a NoCoda violation.

9 Some forms, like the imperative ending /e, ro/, or potential ending /–e, rare/, can in any case only be dealt with as allomorphy.

is now a V-stem, ruling out the V-initial passive allomorph at this stage (*nom-ase-*are*, nom-ase-*rare*). On the other hand, suffixes ending in consonants, such as the alternate causative form /-sas,-as/, make another C-stem (9b), requiring the V-initial allomorph (nom-as-*are*, *nom-as-*rare*) (see Miyagawa 2012 and work cited there for the syntactic differences between the two types of causatives).

(9) a.


b.


The hierarchically organized word structure makes clear that the composition of stem+affix is another (larger) stem, with further affixation following the same requirements, with no special provisos.

It is perhaps useful to contrast this kind of analysis with the approach taken in traditional school grammar (*gakkō bunpō*), based on the grammatical rules proposed by Shinkichi Hashimoto (see Hashimoto 1948). Here three types of verbal conjugations are distinguished: *godan katsuyō* ('5-step conjugation'), *shimo-ichidan katsuyō* ('lower 1-step conjugation' = *e*-row conjugation), and *kami-ichidan katsuyō* ('upper 1-step conjugation' = *i*-row conjugation). The numerals refer to the kana-syllabary arrangement, composed of CV-units in a 5×10 table, where the rows correspond to each of the five vowels of Japanese /a,i,u,e,o/, so that each column is composed of a different consonant /Ca, Ci, Cu, Ce, Co/.[10] The terminology refers to the fact that 5-step verbs (=C-stems) use all five rows of the syllabary in their conjugation (no*ma*nai, no*mi*masu, no*mu*, no*me*ba, no*mo*o 'drink'), whereas the 1-step verbs (=V-stems) use only the *e*-row (to*me*nai, to*me*masu, to*me*ru, to*me*reba, to*me*yoo 'stop') or the *i*-row (si*mi*nai, si*mi*masu, si*mi*ru, si*mi*reba, si*mi*yoo 'soak, permeate'). The terms *shimo-* 'lower' and *kami-* 'upper' refer to the row position in the syllabary (*i* occupying an upper row, *e* occupying a lower row). Since all roots and suffixes must be analyzed into CV-kana units, the roots 'drink' and 'stop' are taken to be /no/ and /to/ (instead of /nom/ and /tom/), and what is actually a root-final C is instead reanalyzed as part of the suffix, i.e., no-*ma*nai, no-*mi*masu, no-*mu*, no-*me*ba, no-*mo*o, and to-*me*nai, to-*me*masu, to-*me*ru, to-*me*reba, to-*me*yoo).[11] An especially odd result, or *reduction*

10 A similar arrangement is found in the Sanskrit syllabary.
11 It is not inconceivable that this approach might have some validity in terms of the way morphological segmentation is actually done by speakers, as in the well-known Maori case analyzed by Hale (1973) in his work on deep-surface disparities, where originally root-final consonants are reanalyzed as suffix-initial, resulting in a system with multiple phonologically unpredictable suffix allomorphs. The issue has remained controversial, however (see McCarthy 1981).

ad absurdum, of this traditional analysis is that short verbs like *miru, minai*, 'see' now have a null root since the initial consonant needs to be part of the suffix. School grammars usually note that these forms do not have a distinction between suffixes and roots, or that they simply do not have roots (in a verbal paradigm chart, the root column is empty; see Suzuki 1972 and Suzuki 1996 for a critical assessment of *gakkō bunpō*).

2.3 The morphophonology of T-suffixes

Different from the C/V-suffixes that have C-initial and V-initial allomorphs, the T-suffixes invariably start with the voiceless coronal plosive or its voiced variant (e.g., /-te, -de/ GERUNDIVE, /-ta, -da/ PAST, /-tara, -dara/ CONDITIONAL). As observed by Bloch (1946a), they do not have V-initial allomorphs and attach directly to both V-stems (tabe-*te* 'eat-gerund', mi-*te* 'see-gerund') and C-stems (tat-*te* 'stand-gerund', sin-*de* 'die-gerund'), the latter resulting in a surface form with a NOCODA violation. The C/V-suffixes can avoid both the ONSET and the NOCODA violation by choosing the appropriate allomorph, but the T-suffixes have no such recourse in their choice of allomorphy. T-suffixation leads to a different kind of allomorphy, as illustrated in (10), namely stem allomorphy, known as the *onbin* (sound change) form.

(10)

verbal stem	*onbin* form	GERUNDIVE	cf. CAUSATIVE-GERUNDIVE
tabe-	–	tabe-te	tabe-sase-te
mi-	–	mi-te	mi-sase-te
hasir-	hasi***t***-	hasit-te	hasir-ase-te
moraw-	mora***t***-	morat-te	moraw-ase-te
nom-	no***n***-	non-de	nom-ase-te
asob-	aso***n***-	ason-de	asob-ase-te
oyog-	oyo***i***-	oyoi-de	oyog-ase-te
kak-	ka***i***-	kai-te	kak-ase-te

Although the sources of the alternations lie in historical sound changes, the phonological shapes of the alternate (*onbin*) stem forms are not accidental. Synchronically, the changes are exactly such that they produce allowed surface codas in Japanese, namely, the first half of a voiceless geminate (here, necessarily *t*, because of the following T-suffix), place-assimilated nasals (here, assimilated to the T-suffix, which itself undergoes postnasal voicing), or a vocoid (the diphthongal off-glide *i*). As pointed out in Ito (1986, 1989), the gemination condition (and related place-assimilated nasal condition) on codas in Japanese turns out to have cross-linguistic and theoretical significance in that it holds with minor variations in many typologically diverse languages, such as Lardil, Diola Fogny, Ponapean, Italian, and Finnish. This common restriction on syllable codas came to be known as CODACOND

in Optimality Theory, and continues to play a prominent role in the theory of positional licensing. T-suffixes attach to the *onbin* form of the stem and not to its basic form because attaching to the latter would lead to violations of CODACOND (e.g., **nom-de, *tob-de, *hasir.te, *kak.te*).[12]

Although the various alternations in the conjugational paradigm appear complex, it is remarkable that once the morphological and allomorphic considerations, including their historical sources, are properly understood, the phonology that is playing a role to derive the surface forms is restricted to the universal syllable structure conditions ONSET, NOCODA, and CODACOND. This result may be surprising, given what appears to be a controversy over the existence of the syllable in Japanese (Labrune 2012). While the importance of the mora in Japanese is well-known and undisputed (see Kubozono 1990, 1995, 1999, and references cited therein; see also Otake, this volume), there is actually also considerable evidence for the syllable as a prosodic constituent in Japanese (see Kawahara 2012 for a summary of acoustic and psycholinguistic evidence). In addition, by eschewing the syllable, proponents of the syllable-less theory must posit different types of moras with different properties, recapitulating syllable theory in a different terminology, but unfortunately within a network of assumptions entirely specific to Japanese. The terminology employed, such as dependent, deficient, or special moras, is reminiscent of how a weak syllable exists only in relation to the strong syllable in the foot. Could it be, then, that a *dependent/special* mora exists only in relation to the *independent/normal* mora in the syllable? Japanese-specific notations such as Q, N, and R depicting moraic obstruents, moraic nasals, and vowel length, respectively, may be useful in transliterating the kana orthography, but do not lead to cross-linguistic phonological understanding or discoveries. Denying the syllable thus comes at a cost: While it is no doubt possible to restate each syllable-based property in roundabout ways that do not refer to the syllable, such an approach would not have allowed Japanese phonological structure to serve as a window through which cross-linguistic syllable conditions could be explored.

3 Phonological alternations

Besides the syllable-conditioned allomorphy discussed above, affixation gives rise to several segmental processes within the syllable (section 3.1) and across syllables (section 3.2), as well as to the formation of superheavy syllables (section 3.3) and accentual alternations (section 3.4). We take up each of these cases, focusing on their

12 Archaic gerundive forms all take the infinitive-*i* form, *nomi-te, tobi-te, hasiri-te, kaki-te*, etc.). S-final stems (*kas-/kasi-*) use the traditional infinitive form ending in *-i* rather than the bare root, and hence do not have an alternate *onbin* form.

cross-linguistic import and their role in the discussion and debates in theoretical phonology.

3.1 Palatalization, affrication, glide deletion

Affixed forms (11) show segmental alternations involving coronal obstruents and high vowels: palatalization and affrication.

(11)

root	/nom/ 'drink'	/kas/ 'lend'	/kat/ 'win'
INFINITIVE	nom-i	kaʃ-i	katʃ-i
PRESENT	nom-u	kas-u	kats-u
IMPERATIVE	nom-e	kas-e	kat-e

Palatalization of coronal obstruents and concomitant affrication of plosives (e.g., *si* → *ʃi*, *ti* → *tʃi*) in the environment of high front vowels is found in many languages, such as Korean, Portuguese, or Mixtec (see Bhat 1978, Bateman 2007, and Kochetov 2011). Palatalization is assimilatory, and affrication of plosives results from the articulatory difficulty of producing a complete oral closure in the palato-alveolar region. While the assimilatory change before *i* is phonetically natural and widely attested cross-linguistically, affrication before *u* (e.g., *tu* → *tsu*) is rare (even though not without parallels: a similar case occurs in Lomongo (Bantu), see Kim 2001), and its causes are not as well-understood. The linguistic term "crazy rule", referring to this process in Japanese, is due to Bach and Harms (1972), an influential early paper in generative phonology which argues that a series of natural sound changes and reasonable innovative generalizations can result in a synchronically phonetically arbitrary "crazy rule." The term has been applied to many phenomena in various other languages, in particular, surrounding the discussion of Evolutionary Phonology (Blevins 2004, etc.). Even though Japanese affrication started out as the original "crazy rule", there is some irony in the fact that it remains an open question whether the process really has no synchronic phonetic (acoustic, articulatory, or perceptual) motivation (see Kim 2001 for an aerodynamic account).

Not as well-known as the palatalization/affrication facts of Japanese, but of potential relevance in this context, is the somewhat odd depalatalization requirement before the mid front vowel *e*, where sequences such as **ʃe,*tʃe*, etc., are dispreferred (and nonexistent in the Yamato and Sino-Japanese strata of the vocabulary).[13] There are no parallel restrictions for back vowels, where both *ʃ* and *s* occur and form a phonemic contrast (*ʃa/sa*, *ʃu/su*, *ʃo/so*). With front vowels, there is no contrast in Yamato and Sino-Japanese: *i*-triggered palatalization leads to *ʃi/*si*, and *e*-triggered

13 See Kubozono (Ch. 8, this volume) and Pintér (this volume) for the appearance of such sequences in the loan vocabulary.

depalatalization to *ʃe/se. In terms of cross-linguistic typology, while it is not surprising that palatalization is restricted to high front vowels, the concomitant depalatalization requirement triggered by mid front vowels is odd (and perhaps qualifies as another "crazy rule": see Ito and Mester 1995, 2003 for a possible approach in terms of contrast; for general discussion of the issue of "crazy rules", see Hyman 2001, Yu 2004, and Scheer, in press).

Finally, due to the phonologically limited distribution of prevocalic glides, syllable-internal glide~Ø alternations are in evidence in the derivational verbal morphophonology. The palatal glide occurs only before back vowels (*ya, yu, yo, *yi, *ye*), leading to *y*~Ø alternations in derivations with verbal roots ending in *y* (such as *moy-* 'burn', or *hay-* 'grow') in combination with transitivizing/intransitivizing affixes (*moy-as-u* vs. *moØ-e-ru*, *hay-as-u* vs. *haØ-e-ru*). The back glide has an even more limited distribution and is found only before the low vowel *a* (i.e., ✓*wa* vs. **wi, *we, *wo, *wu*), so that *w*-final verbal roots only preserve their final glide with *a*-initial suffixes (e.g., *kaw-* 'buy': *kaw-ase* CAUSATIVE, *kaw-are* PASSIVE, *kaw-anai* NEGATIVE PRESENT, but *kaØ-i* INFINITIVE, *kaØ-u* PRESENT, *kaØ-oo* VOLITIONAL, *kaØ-eba* PROVISIONAL).[14]

3.2 Voicing, nasalization, feature preservation

In section 2.3 above, we saw how T-suffixes on C-stems trigger stem allomorphy, so that the resulting *onbin* stem forms are phonologically well-structured (i.e., do not violate CODACOND). Besides the noted changes in the stem, the T-suffix itself is subject to voicing alternations that reveal processes noteworthy in the context of cross-linguistic typology: postnasal voicing, nasalization, and feature preservation.

Because of the universal cross-linguistic dispreference (and avoidance) of NC̥ clusters (=nasal followed by a voiceless obstruent), the T-suffixes in (12) following a nasal(-final) stem appear in their voiced variant (*ʃin-de, kakon-de*), rather than their voiceless variant (**ʃin-te, *kakon-te*).

(12) Postnasal voicing: Nasal-final stems

		n-stem	m-stem	
	root	/sin/ *die*	/sum/ *live*	/kakom/ *surround*
	GERUNDIVE	ʃin-de	sun-de	kakon-de
	PAST	ʃin-da	sun-da	kakon-da
	CONDITIONAL	ʃin-dara	sun-dara	kakon-dara
cf.	*PRESENT*	ʃin-u	sum-u	kakom-u
	PROVISIONAL	ʃin-eba	sum-eba	kakom-eba

14 For the allomorphy analysis mentioned earlier, an interesting question remains as to how the V-initial suffix is chosen for these glide-final roots, given that the glide is absent in the surface form, leading to an ONSET violation. A purely output-oriented allomorphy choice would lead to the wrong C-initial suffix (**ka-ru* instead of *kaØ-u*, etc.).

Postnasal voicing in Japanese played a crucial role in the discussions surrounding underspecification (Ito and Mester 1986; Mester and Ito 1989; Ito, Mester, and Padgett 1995) in that it seemed to constitute a case where a redundant feature specification acted in ways otherwise reserved for distinctive specifications. In illustrating the various ways in which languages choose to avoid NC̥ (13), Pater (1999) gives the Yamato stratum of Japanese as a prime example of one of the strategies of NC̥ avoidance, namely postnasal voicing (e.g., *nt* → *nd*).

(13) NC̥ avoidance strategies

nt →	nd	Yamato Japanese, Puyo Pungo Quechua
	tt	Mandar
	nn	Konjo
	n	Umbundu
	t	Kelantan Malay

Interestingly, not every conceivable way of resolving the NC̥ problem is actually found in natural languages. Thus vowel epenthesis is apparently never used in this context, according to Pater – one of the first illustrations of what later came to be known as the "too-many-solutions problem" in OT (Steriade 2001).

Related to postnasal voicing is the nasalization of voiced-obstruent-final stems in (14).

(14) Nasalization: Voiced obstruent final stems

		b-stem	
	root	/asob/ 'play'	/narab/ 'line up'
	GERUNDIVE	ason-de	naran-de
	PAST	ason-da	naran-da
	CONDITIONAL	ason-dara	naran-dara
cf.	PRESENT	asob-u	narab-u
	PROVISIONAL	asob-eba	narab-eba

Because of CODACOND, place assimilation/gemination occurs at the stem-suffix boundary (see section 2.3 above for examples). With a stem-final voiced obstruent, however, we find nasalization of the place-assimilated coda (*ason-de, naran-de*) rather than the voiced obstruent geminate (**asod-de, *narad-de*). Another general cross-linguistic constraint is at work here, this time, against voiced obstruent geminates (for a recent study, see Kawahara 2006, Kawahara, Ch. 1, this volume, and Kawagoe, this volume).

Finally, the voiced variant of the T-suffix occurs with stems ending in the voiced velar obstruent *g*. This would be the expected variant, if it were not for the fact that there is no stem-final trace of this obstruent voicing when the T-suffix is attached

since both *k*-final and *g*-final stems themselves occur in *onbin* forms ending in *i* ("velar vocalization").[15]

(15) Vocalization and Feature Preservation: Velar-final stems

		k-stem		*g*-stem	
	ROOT	hik- *pull*	kawak- *dry*	tog- *sharpen*	tsug- *pour*
	GERUNDIVE	hii-te	kawai-te	toi-de	tsui-de
	PAST	hii-ta	kawai-ta	toi-da	tsui-da
	CONDITIONAL	hii-tara	kawai-tara	toi-dara	tsui-dara
cf.	PRESENT	hik-u	kawak-u	tog-u	tsug-u
	PROVISIONAL	hik-eba	kawak-eba	tog-eba	tsug-eba

Since the voicing of the suffix can only be determined by considering the feature composition of the input, this opaque alternation has attracted the attention of optimality-theoretic analysts. Lombardi (1998) argued that it constitutes strong evidence for feature-level faithfulness (MAXFEATURE), where the input voicing feature in the input is preserved by docking onto another segment in the output. OT analyses sometimes employ such feature-faithfulness constraints (MAXFEATURE, DEPFEATURE), but most cases can be restated in terms of segmental faithfulness (IDENTFEATURE) constraints. However, as convincingly shown by Lombardi, the voicing triggered by underlying stem-final *g* in Japanese cannot be reduced to segmental faithfulness because the relevant input segment carrying the voicing feature in question arguably corresponds to the stem-final *i*, and not to the voiced suffix-initial segment *d*. The T-suffix alternation with *g*-final stems thus remains one of the few convincing cases of feature-level faithfulness in OT.

3.3 Superheavy syllables

Universal syllable theory countenances, besides light (monomoraic) and heavy (bimoraic) syllables, superheavy syllables (trimoraic, or even heavier) as a marked option. The syllable inventory of Japanese is no exception: The overwhelming majority of words are composed of maximally bimoraic syllables. Most of the superheavy syllables are found in the peripheral (recent Western) loan vocabulary, such as *toon* 'tone', *pataan* 'pattern', *sain* 'sign', or *toronboon* 'trombone'. In the core vocabulary, Sino-Japanese items admit no superheavy syllables, reflecting the size restriction

15 Historically the result of intervocalic lenition of the velar stop: *aki* > *axi*> *ai*, *agi* > *aɣi* > *ai*. Velar vocalization is also found in the adjectival conjugation (archaic *waka-ki*, contemporary *waka-i* 'young').

allowing only maximally bimoraic morphemes (see Tateishi 1990, Ito and Mester 1996, and Ito and Mester, Ch. 7, this volume). In native Yamato items, on the other hand, while superheavy syllables are not found in underived forms (morpheme-internally), what appear to be superheavy syllables come about as a result of morpheme concatenation and derivation. One such case arises through the by-now-familiar affixation of T-suffixes when they attach to verbal roots of the shape /CVVC-/, such as *toor-* and *hair-*, as shown in the paradigm in (16).

(16)

		'pass'	'freeze'	'enter'	'come, visit'		'take'	'paste'
	root	toor-	koor-	hair-	mair-	*cf.*	tor-	har-
	GERUNDIVE	toot.te	koot.te	hait.te	mait.te		tot.te	hat.te
	PAST	toot.ta	koot.ta	hait.ta	mait.ta		tot.ta	hat.ta
	CONDITIONAL	toot.ta.ra	koot.ta.ra	hait.ta.ra	mait.ta.ra		tot.ta.ra	hat.ta.ra
cf.	PRESENT	too.ru	koo.ru	hai.ru	mai.ru		to.ru	ha.ru
	PROVISIONAL	too.re.ba	koo.re.ba	hai.re.ba	mai.re.ba		to.re.ba	ha.re.ba

Here C/V suffixes (such as the present and provisional forms shown above) choose their V-initial allomorphs (*-u*, *-eba*), so the last consonant in the verbal root becomes an onset (too.*ru*, hai.*ru*), and the remainder is a bimoraic syllable (*too*.ru, *hai*.ru). T-suffixes, however, have no V-initial allomorph, and the verbal root in its *onbin* form is syllabified into one syllable (*toot*.te, *hait*.te). These forms must be analyzed as containing superheavy syllables (*to'ot*.te, *ha'it*.te, etc.) accented on their first mora (*o* and *a*). If the apparent superheavies were split into two syllables (*to.*o't*.te, *ha.*i't*.te), the accent would be predicted to fall on the mora/syllable immediately preceding the T-suffix, like ha.*ʃi't*.te 'run-GERUNDIVE'. There is no tendency to shorten the vowels of superheavies like *tootte*, which remains distinct from *totte* 'take-GERUNDIVE' in all styles of speech (Vance 2008).

Other apparent instances of superheavies are less clear, and the status of such trimoraic syllables in Japanese has remained controversial. Several researchers (Kubozono 1999: 50–55; Vance 2008: 132) have provided arguments that some (or all) of the purportedly trimoraic syllables should be analyzed as broken into two syllables (monomoraic + bimoraic). The evidence comes from native intuition for syllable boundaries, the possibility of vowel rearticulation, and, most convincingly, from patterns of accentuation, requiring further investigation and analysis. Controversial cases arise with derivational (adjective- and noun-forming) suffixes that are geminate-initial, such as the denominal adjective-forming suffix *-ppoi* '-ish' and the suffix *-kko* (lit. 'child/person').

(17)

	-ppoi	'-ish, -like'
kodomo	ko.do.mop.poi	'childish, childlike'
onna	on.nap.poi	'womanly'
tihoo	ti.hoop.poi	'country-like'
sutaa	su.taap.poi	'(pop-)star-like'
doraemon	do.ra.e.monp.poi	'like Doraemon (cartoon figure)'
ebisen	e.bi.senp.poi	'like shrimp-flavored rice cracker'

	-kko	'person from' (*demonym)*
Edo	e.dok.ko	'Edo-ite'
Sendai	sen.daik.ko	'Sendai-ite'
Pari	pa.rik.ko	'Parisian'
Rondon	ron.donk.ko	'Londoner'
Berurin	be.ru.rink.ko	'Berliner'
Uiin	u.iink.ko	'Wiener'

Here the accentual evidence is conflicting: While *sendai'kko* (**senda'ikko*) suggests that the apparent superheavy is broken into two syllables (*sen.da.i'k.ko*), this is not persuasive for *rondon'kko* (**rondo'nkko*), which is unlikely to be syllabified as *ron.do.n'k.ko*, given that a syllable *.nk.* is otherwise unheard of in Japanese in any context. What cases like these suggest is that, rather than trying to analyze superheavies away, we need to better understand the way they behave with respect to accent rules.

Since the suffixes *-ppoi* and *-kko* are phonologically unrestricted in their combinatorics, the last syllable of the stem can become heavy (*ko.do.mop.poi*, *e.dok.ko*, etc.) or superheavy (*ti.hoop.poi*, *ron.donk.ko*, etc.) upon suffixation (i.e., modulo the reservations noted above). They can even attach to a superheavy syllable (u.*iin* 'Wien'), resulting in an apparent ultra-superheavy tetramoraic syllable u.*iink*.ko (17).[16] Suffixation of these geminate-initial suffixes is productive, indicating that such superheavy syllables are allowed as marked options at morphological junctures.

3.4 Suprasegmental properties of affixation

Besides their various phonological and allomorphic differences in segmental composition discussed in the previous sections, verbal (and adjectival) stems are morphophonemically characterized by a suprasegmental property of the stem, a tonal fall (analyzed as [±accent] by McCawley 1968). The [±accent] feature is an underlying suprasegmental property of the verb (i.e., just like any segmental property, it

16 Similar formations are found with the quotative suffix *-tte*, which freely attaches to syllables of all kinds, including superheavies (*u.iint.-te itta* 'Vienna, (s)he said').

is not predictable whether a certain verb is marked as [+accent] or [–accent]). On the other hand, different from segmental features like [labial], the [accent] feature is not inextricably docked onto a particular segment. As illustrated in (18), the accent (marked by an apostrophe indicating the tonal fall) migrates towards the end of the entire stem complex and ends up appearing on suffixes, not on the root that sponsors it. In the [–accent] column, there is no accentual fall in the verbal stem complex.

(18)	[+accent]		[–accent]	
ROOT	sirabe- 'investigate'	tanom- 'request'	narabe- 'line up'	susum- 'advance'
PRESENT	[sirabe']-ru	[tano'm]-u	[narabe]-ru	[susum]-u
CAUS-PRES	[sirabe-sase']-ru	[tanom-ase']-ru	[narabe-sase]-ru	[susum-ase]-ru
PASS-PRES	[sirabe-rare']-ru	[tanom-are']ru	[narabe-rare]-ru	[susum-are]-ru
CAUS-PASS-PRES	[sirabe-sase-rare']-ru	[tanom-ase-rare']-ru	[narabe-sase-rare]-ru	[susum-ase-rare]-ru
IMPERATIVE	[sirabe']-ro	[tano'm]-e	[narabe]-ro	[susum]-e
PROVISIONAL[17]	[sirabe']-reba	[tano'm]-eba	[narabe]-re'ba	[susum]-e'ba

For T-suffixes (19), the accent is placed on the *penultimate* mora of the stem to which the T-suffix attaches (compare (18), where the accent appears on the *final* mora of the relevant stem).

(19)	[+accent]		[–accent]	
PAST	[sira'be]-ta	[tano'n]-da	[narabe]-ta	[susun]-da
CAUS-PAST	[sirabe-sa'se]-ta	[tanom-a'se]-ta	[narabe-sase]-ta	[susum-ase]-ta
CAUS-PASS-PAST	[sirabe-sase-ra're]-ta	[tanom-ase-ra're]-ta	[narabe-sase-rare]-ta	[susum-ase-rare]-ta
GERUND	[sira'be]-te	[tano'n]-de	[narabe]-te	[susun]-de
CONDITIONAL	[sira'be]-tara	[tano'n]-dara	[narabe]-ta'ra	[susun]-da'ra
ALTERNATIVE	[sira'be]-tari	[tano'n]-dari	[narabe]-ta'ri	[susun]-da'ri

The difference in accentual behavior between T-suffixes and C/V-suffixes might be due to the fact that T-suffixes prosodically subcategorize for a bimoraic foot (see section 5.1 below), thereby attracting the underlying accent to the head of this foot: *si(ra'be)-ta, sirabe(sa'se)ta, ta(no'n)-da* (for discussion of prosodic subcategorization, see Inkelas 1989, McCarthy and Prince 1990b, and Paster 2006). C/V suffixes, on the other hand, form a foot with the last syllable of the stem they attach to, attracting the underlying accent to this position.[18]

17 The disyllabic suffixes, *-re'ba* (18) as well as *-ta'ra, -ta'ri* in (19), are underlyingly accented, but their accents are deleted when preceded by accented stems: *sirabe'-re'ba* → *sirabe'reba*. The deletion is usually analyzed as triggered by a type of OCP violation. According to recent instrumental studies (such as Kubozono 1988/1993), however, in careful pronunciation it is possible for both accents to be realized in *sira'be-ta'ri*, but not in *sirabe'-re'ba* with accents on adjacent syllables (which conceivably incur a more serious OCP violation). If each accent is a bitonal HL complex, tonal overcrowding may also be a factor leading to simplification, a topic for future research.
18 The distinction in accent location between the two kinds of suffixes is neutralized when the syllable preceding a T-suffix is heavy (*tano'mu, tano'nda* 'request, requested').

These accentual properties, as well as the segmental alternations discussed in the previous sections, carry over to newly formed verbs with stem-forming *-r-* (*ku'mo* 'cloud', *kumo'r(u)* 'to cloud', see section 2.1 above). Alongside the compound structure *X-suru* 'do X' with the light verb *-suru*, where X is a loanword, we also find *X-r(u)* formations, as in (20).

(20)		me'mo-suru 'to jot down a memo'	mi'su-suru 'to miss, make a mistake'	ko'pii-suru 'to copy, do copying'
a.	PRES	[memo'r]u	[misu'r]u	[kopi'r]u
b.	CAUS-PRES	[memor-ase']ru	[misur-ase']ru	[kopir-ase']ru
c.	CAUS-PAST	[memor-a'se]ta	[misur-a'se]ta	[kopir-a'se]ta
d.	GERUND	[memo't]te	[misu't]te	[kopi't]te

The accentuation is as expected: final accent in the relevant stem complex with C/V suffixes (20a,b), and penultimate accent in the same domain with T-suffixes (20c,d).

Stem forming *-r-* is not as productive as *-suru* compounding, so many loanwords do not have the *r*-form. Loanwords longer than two moras do not generally seem to acquire the *r*-suffix.[19]

(21)	dora'ibu –suru	'to go for a drive'	*doraibu'-r(u)
	kome'nto- suru	'to make a comment'	*komento'-r(u)
	hi'tto-suru	'to hit, to make a hit'	*hitto'-r(u)

The form *kopi-r(u)* cited above show shortening of its final vowel to make the stem form fit the bimoraic template. Other shortened *r*-verbs include *negu-r(u)* 'to neglect' and *sabo-r(u)* 'to sabotage, skip classes'. In computerese, *r*-forms are more prevalent and apparently do not require the stem to be bimoraic (22).

(22)	ha'ngu-suru	'to hang'	hangu'ru	hangu'tta	'to hang (as computer problem)'
	huri'izu-suru	'to freeze'	huriizu'ru	huriizu'tta	'to freeze (as computer problem)'

When there is a final syllabic /l̩/ in the source word (borrowed as /ru/ in the loanword) (23), as in *daburu* 'double', *toraburu* 'trouble', and *guuguru* 'Google', /ru/ is able to do double duty as the original source word ending and the stem-forming affix.

19 This is an example of templatic word formation, in the sense of section 5 below. New adjectives involving truncation also belong in this category *kisyo-i* < *kisyoku-waru-i* 'disgusting', *kimo-i* < *kimoti-waru-i* 'unpleasant', with a trimoraic option, as in *mendo-i* < *mendo-kusa-i*, 'troublesome', *utto-i* < *uttoosi-i* 'gloomy, unpleasant'.

(23)	da'buru	'double'	dabu'ru	dabu'tta	'to double' PRESENT, PAST
	tora'buru	'trouble'	torabu'ru	torabu'tta	'to trouble' PRESENT, PAST
	guuguru	'Google'	guugu'ru	guugu'tta	'to Google' PRESENT, PAST

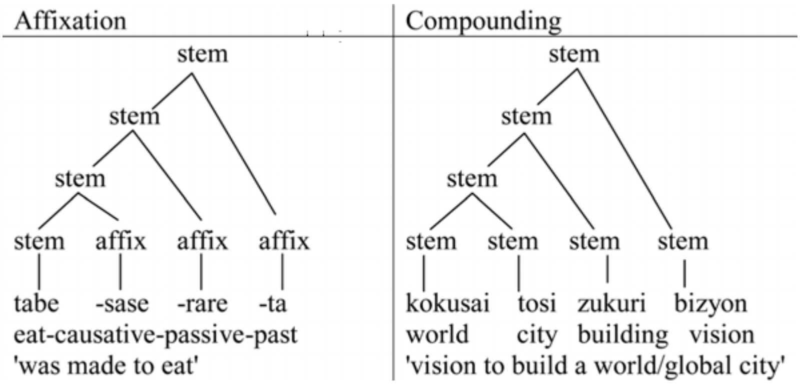
These novel *ru*-formations show that the verbal accentuation pattern is exceptionless and overrides the original noun accent, as in *da'buru* (noun) vs. *dabu'ru* (verb).

While morphophonemic properties as exemplified in Japanese affixational morphology are sometimes dismissed as idiosyncratic, or as mere historical residues, even a cursory look at the segmental, prosodic and accentual phenomena involved, some of them quite productive, reveals many interesting phonological processes and properties of Japanese, with cross-linguistic consequences that still await closer investigation.

4 Phonology of compounding

Besides affixation, which joins an affix to a stem to form another stem, the other major word formation process in Japanese is compounding, which joins one stem to another, as exemplified in (24), repeated from (1).

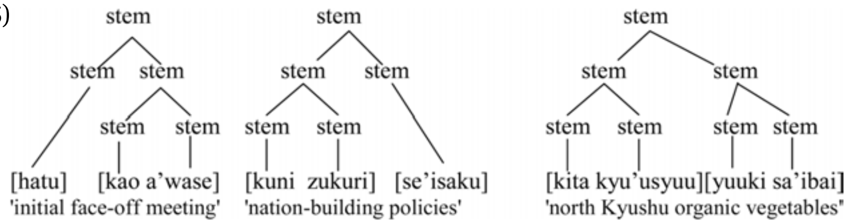
(24)



Affixation, predominantly suffixal in Japanese,[20] usually results in left-branching recursive word structure as in (24a), whereas a variety of recursive structures, with other kinds of branching, can be found in compounding (25).

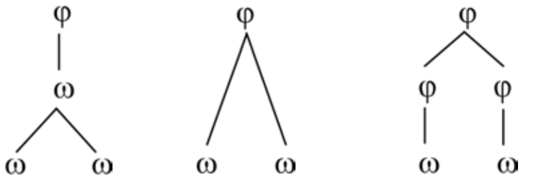
20 Some morphemes usually analyzed as prefixes have interesting phonological properties, e.g., the honorific prefixes *go-/o-*, or the phrasal prefixes *moto-/kyuu-*, 'former', *datu-* 'escaping', etc. Whether these cases are to be properly analyzed as prefixation or as a type of compounding remains a question for further exploration (see Irwin 2012).

(25)



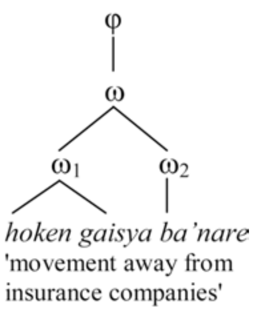
While these compounds behave syntactically as single lexical words and not as phrases (i.e., case marking only on the final stem, no optional scrambling, no internal modifiers), it turns out that they have different phonological characteristics, such as *rendaku* voicing (see Vance, this volume, and Ito and Mester 2003), junctural accent, and deaccenting (see Kawahara, Ch. 11, this volume). Building on Kubozono, Ito, and Mester (1997), Ito and Mester (2007) propose that these differences can be understood as the result of the particular way the syntactic/morphological structure is mapped to prosodic structure (see Ishihara, this volume), crucially allowing for recursion both at the word (ω) and phrase (φ) levels. Depending on various factors, a lexical compound ends up being phonologically parsed as a word compound (26a), as a phrasal compound with two prosodic words (26b), or as a phrasal compound with two phonological phrases (26c).

(26) a. word compound b. mono-phrasal compound c. bi-phrasal compound



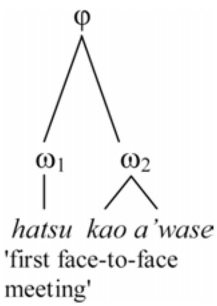
With further compounding, a number of structural possibilities are encountered, as shown in the examples in (27) taken from Ito and Mester (2007), where the lowest branching ω's are themselves word compounds.

(27) a.

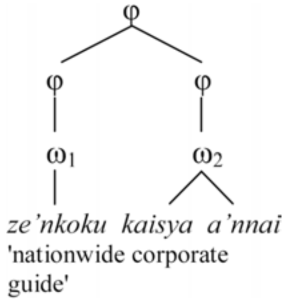


b.

c.



d.



These complex compounds exhibit an interesting distribution of phonological properties. In (27a), ω_2 undergoes *rendaku* voicing (*hanare* → ***b**anare*); but in (27b–d), ω_2 does not undergo *rendaku*, although the individual stems are *rendaku* undergoers in simple compounds such as *kara-**b**uri* 'empty swing', *e-**g**ao* 'smiling face', and *takusii-**g**aisya* 'taxi company'. Junctural accent (here realized on the initial syllable of ω_2, but see Kawahara, Ch. 11, this volume, for details) is found in (27a,b) but not in (27c,d), where it is precluded by the phonological length of ω_2. Finally, because of the one-accent restriction on φ, there is no accent on ω_1 in (27a–c), whereas in (27d), each ω constitutes a separate φ, and therefore ω_1 carries its own accent. The overall pattern is summarized in (28), where we can see a clear progression with the word compound (27a) exhibiting all three properties, and the biphrasal compound (27d) exhibiting none of them.

(28)

	(27a)	(27b)	(27c)	(27d)
Rendaku voicing on ω_2	*Yes*	*No*	*No*	*No*
Junctural accent on ω_2	*Yes*	*Yes*	*No*	*No*
Deaccenting of ω_1	*Yes*	*Yes*	*Yes*	*No*

We have here introduced these phonological properties of compounds in only the broadest outline, as it is beyond the scope of this general word formation chapter to present or discuss the details (see Kawahara, Ch. 11, this volume, Vance, this volume, and references cited there).

Finally, not mentioned above, but no less pervasive in the language, is the compounding of Sino-Japanese morphemes, which is associated with a special segmental phonology and prosodic morphology, giving rise to very systematic alternations, such as vowel~Ø-alternations (*dai-gak**u*** 'lit. large-scholarship, university' vs. *gak-koo* 'lit. scholarship-building, school') and *h~p* alternations (*sip-**p**itu* 'lit. take-brush, to write', *toku-**h**itu* 'lit. special-brush, special note', *en-**p**itu* 'lit. led-brush, pencil'

*mannen-**h**itu* 'lit. ten-thousand-year brush, fountain pen'). Sino-Japanese compounding obeys very rigid restrictions on size and segmental combinatorics unknown to the rest of the lexicon (see Tateishi 1990, Ito and Mester 1996, Kurisu 2000, and Ito and Mester, Ch. 7, this volume).

5 Templatic word formation

So far, this chapter has focused on the phonological processes that accompany word formation, taking place when morphemes with a fixed shape are combined in various ways through affixation and compounding. In what follows, we turn our attention to what has become known as prosodic morphology since the seminal work of McCarthy and Prince (1986). Here phonology is not just an accompaniment to word formation, but takes center stage in the process itself by determining the shape of words through phonological (prosodic) templates.

Phonological representation does not consist of just a sequence of vowel and consonant phonemes, but such phonemic segments are organized into a constituent structure known as the prosodic hierarchy (Selkirk 1978; Nespor and Vogel 1983), which is not isomorphic to the grammatical (syntactic/morphological) hierarchy.

(29)

Linguistic hierarchies	Grammatical hierarchy	Prosodic hierarchy
phrase-level hierarchies	sentence	υ: utterance
	|	|
	syntactic phrase	ι: intonational phrase
	|	|
		φ: phonological phrase
		|
word-(internal) hierarchies	morphological word	ω: prosodic word
	|	|
	stem	f: foot
	|	|
	root/affix	σ: syllable
		|
		μ: mora

Phrasal syntactic constituents are mapped to intonational and phonological phrases, and morphological words, formed by affixation or compounding as discussed in the previous sections, are mapped to prosodic words that are internally organized into feet, syllables, and moras. In templatic word formation, such as truncation, there is no affixation or compounding, but the word shape, in particular its size, is phonologically determined by the prosodic template, which is itself dictated by independent phonological constraints.

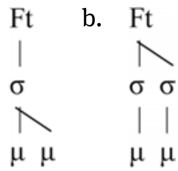
For example, English clipped words such as *rep* for *representative* or *prof* for *professor* are not combinations of preexisting morphemes, but rather the phonology provides the prosodic shape of the clipped word, here, a heavy monosyllable. As McCarthy and Prince (1986) convincingly show, this template corresponds to the minimal prosodic word of the language (see also Lappe 2003) and constitutes the template to which the segments of the base word are mapped, resulting in formations whose prosody is invariant but whose segmentalism varies with their bases.

Templatic word formation has been investigated extensively in Japanese (Poser 1990; Ito 1990; Mester 1990; Ito and Mester 2003 [1992]; Kubozono 1995; Labrune 2002; Benua 1995), providing numerous studies on a wide-ranging variety of formations, such as clipped words, hypocoristics and nickname formation, reduplication, language games, alphabetic acronyms, baseball chants, blends, and compound truncations. Many of these studies have garnered interest not only from Japanese specialists but also from the field of phonology at large. Examples and analyses of Japanese prosodic morphological formations are routinely cited in textbooks (Kenstowicz 1994: 651–653), and in work dealing with templatic word formations.[21] Since it will not be possible here to survey all of these cases, we will exemplify in some detail those formations whose investigation has had an influence on the direction of phonological theorizing, in particular, in the developments leading to Optimality Theory.

5.1 The bimoraic foot template

Besides evidence from accentual patterns in compounds, Poser (1984a,b, 1990) presented extensive evidence that the bimoraic foot plays a pivotal role in Japanese word formation, in particular, in hypocoristic and other nickname formations. The bimoraic foot template $_F(\mu\mu)$ subsumes a single heavy (bimoraic) syllable and two light (monomoraic) syllables, as schematically shown in (30).

(30) Bimoraic foot structure: a. b.



21 Although not directly templatic, Kubozono (1990) shows that language-specific prosodic properties hold the key to the different generalizations behind the process of blending word formation in English and Japanese, where the two merged words are in a paradigmatic relation (i.e., *smog* (*smoke+fog*) is neither smoke-like fog nor fog-like smoke, but a genuine mixture of the two). Similarly, the Japanese blend *gozira* ‘Godzilla’ is a mix of *gorira* ‘gorilla’ and *kuzira* ‘whale’. Following Kubozono’s work, Bat-El (1996) further develops the phonological basis of blending types with evidence from Hebrew and other languages.

Hypocoristic names (31) are derived by adding the suffix /-tyan/ [tʃaɴ] (besides others), with various segmental alterations to the original (base) name.

(31)	megumi	'Megumi'	>	megu-tyan
	keeko	'Keiko'	>	kee-tyan, keko-tyan
	hiromi	'Hiromi'	>	hiro-tyan, romi-tyan
	midori	'Midori'	>	mido-tyan, mii-tyan
	yooko	'Yoko'	>	yoko-tyan, yoo-tyan
	mariko	'Mariko'	>	mari-tyan, mako-tyan
	hanako	'Hanako'	>	hana-tyan, haa-tyan
	takako	'Takako'	>	taka-tyan, taa-tyan
	akira	'Akira'	>	aki-tyan, at-tyan
	tatuo	'Tatsuo'	>	tat-tyan
	kentaroo	'Kentaro'	>	ken-tyan, taro-tyan
	wasaburoo	'Wasaburo'	>	wasa-tyan, sabu-tyan

It is typical to find several possible hypocoristic forms for a specific personal name (as in English, with *Liz, Lisa, Eli* for *Elizabeth*, or *Al, Albie, Bert* for *Albert*). The possible variations in Japanese hypocoristics are segmental deletions at the edges of names (*hiro-tyan, romi-tyan* from *hiromi*), segment skipping (*mako-tyan* from *mariko*), and, in combination with segment deletion, vowel lengthening (*mii-tyan* from *midori*), vowel shortening (*keko-tyan* from *keeko*), and gemination from the suffix (*at-tyan* from *akira*). What unifies the hypocoristic name formation, Poser argues, is the overall size demand, namely a bimoraic foot template that is filled by the segments of the base. From the name *midori*, we find bimoraic *mido-tyan* or *mii-tyan*, but not a monomoraic **mi-tyan*.

Other types of truncatory names are also discussed by Poser (1990), such as the rustic girl's names (*keeko>o-kee, takako>o-taka*) or discretionary names of clients (*koono>o-koo-san, tanizaki>o-taa-san*), which have different affixal attachments and stricter segmental requirements, but the bimoraic foot template of the modified base continues to be invariant (see Mester 1990 for discussion and analysis in terms of Prosodic Morphology).

Established long compounds often have abbreviated forms that also conform to this size restriction, taking the initial foot-sized portion of each compound member (32a,b), or taking only the initial (modifying) word of the compound (32c,d). Similar abbreviation strategies also exist in English, such as taking the initial letter of each word to form an acronym (TEPCO, PC), or abbreviating the compound to its first word, as in *cell* for *cell phone*, or *super* for *superintendent*.[22] What is remarkable for

22 Different from Japanese, English *super* is an abbreviation for *superintendent* or *supernumerary* 'extra (actor in film)', not for *supermarket*.

Japanese is the strict prosodic generalization that the abbreviated result is maximally two bimoraic feet which do not necessarily match with any morphological divisions, as shown by *(zi+te)n+* in (32e) and *(ho+ko)o+* in (32f).

(32)	Compound abbreviations			$_F(\mu\mu)$ $_F(\mu\mu)$	
	a.	too+kyoo den+ryoku	→	(too)(den)	'Tokyo Electric Power Co., TEPCO'
	b.	paasonaru konpyuutaa	→	(paso)(kon)	'personal computer, PC'
	c.	kee+tai den+wa	→	(kee)(tai)	'mobile/cell (phone)'
	d.	suupaa-maaketto	→	(suu)(paa)	'supermarket'
	e.	*zi*+*te*n+sya tuu+kin	→	(zite)(tuu)	'bicycle commuting'
	f.	*ho*+*ko*o+sya ten+goku	→	(hoko)(ten)	'pedestrian paradise, no-car zone'

The bimoraic foot also serves as a size restriction for Sino-Japanese (bound) morphemes, e.g., /gen/ 'speak', /go/ 'word', /gaku/ 'scholarship, combining to form words like *gengo* 'language' and *gengogaku* 'study of language, linguistics' (see Ito and Mester, Ch. 7, this volume), and for so-called mimetics (sound-symbolic morphemes usually occurring reduplicated, e.g., *kon-kon* 'knocking' or *pota-pota* 'dripping'; see Nasu, this volume).

Different from (reduplicated) mimetics, verbal reduplication prima facie shows no evidence of a foot-size restriction. The stem form of the verb, however long, is reduplicated in its entirety, deriving an adverbial form meaning 'while V-ing' (33).

(33)	tabe	'to eat'	tabe-tabe	'while eating'
	naki	'to cry'	naki-naki	'while crying'
	odori	'to dance'	odori-odori	'while dancing'
	hataraki	'to work'	hataraki-hataraki	'while working'

Even here, though, the foot-size restriction is lurking in the background and manifests itself this time as a minimality condition: When monomoraic verbal stems like *mi* 'see' in (34) reduplicate, both the reduplicated portion and the monomoraic base must lengthen to bimoraic size (*mii-mii*, etc., see Martin 1975: 409, and Kageyama 1976–77: 127).

(34)	mi	'to look'	mii-mii	'while looking'
	ne	'to sleep'	nee-nee	'while sleeping'
	si	'to do'	sii-sii	'while doing'
	benkyoo-si	'to study-do'	benkyoo-sii-sii	'while studying'

Besides providing another example of the bimoraic foot at work in word formation, monomoraic lengthening in verbal reduplication provides an example of what has become known as "backcopying" in the Generalized Template Theory of reduplication developed by McCarthy and Prince (1999). The idea is the following. The base

verb *tabe-* is copied, to form *tabe-tabe*. For *mi-*, however, copying the base form would yield **mi-mi*, which does not satisfy the foot size requirement of the reduplicated portion. Lengthening just the reduplicated portion to satisfy the requirement would yield **mi-mii* or **mii-mi*, depending on whether the reduplicant is suffixed or prefixed. There is clearly no requirement that the base itself has to be minimally bimoraic, otherwise monomoraic verbs such as those in (34) could not exist. However, reduplication requires that the base and its copy should be identical. Therefore the length of the reduplicant is back-copied into the base to achieve identity, resulting in the doubly-lengthened *mii-mii*.

The significance of this back-copying case is that this small corner of Japanese templatic word formation joins a growing number of cross-linguistically attested cases of backcopying (Inkelas and Zoll 2005 for English, Downing 2000 for the Bantu language Kinande, and Caballero 2006 for the Uto-Aztecan language Guarijío), together undermining some of the main hypotheses of Generalized Template Theory (McCarthy and Prince 1999), as argued in detail in Gouskova (2007) (a similar conclusion is reached in McCarthy 2008: 297 and McCarthy, Kimper, and Mullin 2012: 210–211).

5.2 The optimal prosodic word

With Poser's (1984a,b, 1990) convincing demonstration that the bimoraic foot Fμμ is an important structural unit in Japanese, subsequent research found confirmation that F*-templates, where F* denotes integer multiples of bimoraic feet, indeed played a key role in many areas of prosodic morphology (Kubozono 1995; Tateishi 1989; McCarthy and Prince 1990a,b, among many others). Further investigation, however, also revealed the existence of templatic word formations involving a number of other highly systematic prosodic properties beyond the use of F*-templates. Based on the empirical findings of Ito (1990) in prosodic morphology terms, Ito and Mester (2003 [1992]) argue that word clippings, a productive word formation pattern of contemporary Japanese, do not conform to F*-templates per se, but in fact reveal a deeper generalization, namely, the emergence of the optimal (or unmarked) structure of prosodic words in Japanese.

Such truncated words, often involving long loanwords, appear in three prosodic shapes, two of which are the familiar F*-templates (singly or doubly footed structures), and the third consisting of a foot and a light (monomoraic) syllable.

(35)	a. Fμμ	[(suto)] ~~raiki~~	'strike'
		[(demo)] ~~nsutoreesyon~~	'demonstration'
		[(roke)] ~~esyon~~	'(film shooting) location'
	b. Fμμ Fμμ	[(riha) (biri)] ~~teesyon~~	'rehabilitation'
		[(kon)(bini)] ~~ensu sutoa~~	'convenience store'
		[(asu)(para)] ~~gasu~~	'asparagus'

c. $F_{\mu\mu}\ \sigma_{\mu}$	[(ani) me] ~~esyon~~	'animation'
	[(dai) ya] ~~mondo~~	'diamond'
	[(paa)ma] ~~nento~~	'perm(anent wave)'
	[(kon)po] ~~onento~~	'(stereo) component'

Why do we find these three word-clipping patterns (henceforth referred to as F-words (35a), FF-words (35b), and Fσ-words (35c))? The final syllable in Fσ-words is necessarily monomoraic, since a heavy syllable would project a foot and count as the second F in an FF-word. If it was merely the case that there exists one more template for word clippings, Fσ, in some sense intermediate in size between F and FF, this would be descriptively interesting, but of no theoretical import. The question to ask, rather, is why, given the possibility of Fσ-words, there are no σF- words like *[*de*(*mon*)] ~~*sutoreesyon*~~ or *[*ro*(*kee*)] ~~*syon*~~, where the foot is located at the right edge and not at the left edge. According to Ito and Mester (2003 [1992]), the answer is not to be found in another restriction on feet, but rather on an edge-based restriction on prosodic words, favoring left edges of words to be properly footed, a situation found cross-linguistically (known as the initial dactyl effect). This explains why [(*kon*)*po*] is well-formed, but *[*de*(*mon*)] is not.

This left-edge matching requirement, which played an important role in Generalized Alignment Theory (McCarthy and Prince 1993),[23] is not an isolated phenomenon in clipped words, but is found in another corner of Japanese prosodic morphology, namely, in the word-reversing language game *zuuzya-go* (ZG). Here regular words are split in two and reversed (for details and analysis, see Tateishi 1989, Ito, Kitagawa, and Mester 1996, and Sanders 1999), so that the resulting ZG word starts with a foot: *karaoke* → [(*oke*)(*kara*)] 'karaoke', *kusuri* → [(*suri*)*ku*] 'drug', *kaban* → [(*ban*)*ka*] 'bag'. For words of the prosodic shape Fσ ([(*pan*)*tsu*] 'pants', [(*koo*)*ra*] '(Coca) Cola', simple reversal leads to the ill-formed σF, *[*tsu*(*pan*)], *[*ra*(*koo*)], and in just this situation, the game allows for further modification to provide the prosodic word with a left-aligned foot, [(*tsuu*)(*pan*)] or [(*tsun*)*pa*], [(*raa*)(*koo*)] or [(*raa*)*ko*]. Kubozono (2003) points out that the same asymmetry is already present in Japanese baby words (*baaba*, **babaa*).

The explanation turned out to also have theoretical consequences with respect to the interpretation of the Strict Layering principle of the prosodic hierarchy, since it was crucial to be able to distinguish syllables that are parsed into feet from those that were not, paving the way for a more nuanced interpretation of Strict Layering as an optimal, ideal prosodic state rather than an absolute requirement (Prince and Smolensky (1993 [2004]); Selkirk 1996).

23 A similar proposal has been made more recently by Selkirk (2011), who argues for the optimality-theoretic constraint StrongStart, reflecting a universal preference for prominent constituents to be initial in any level of the prosodic hierarchy.

Finally, a closer look at the three possible word types in (35) raises other questions. First, whereas the feet in FF- and Fσ- words consist of either one heavy syllable or two light syllables, F-words are always disyllabic (*suto*, *demo*, etc.), heavy monosyllables (**dai*, **paa*, **kon*, etc.) are never found. Why would this be the case, given that both are licit bimoraic foot structures? Second, given FF-words (conforming to the F*-template), why are there no FFF-words like *[(*kon*)(*bini*)(*en*)]~~*su sutoa*~~? Similarly, given F-, FF- and Fσ-, why no FFσ-words like *[(*asu*)(*para*)*ga*]su, or FσF-words like *[(*kon*)*bi*(*nee*)]~~*syon*~~? Once again, the answer lies not in a constraint on feet, but in a constraint imposed on the overall structure of prosodic words, namely, the Word Binarity constraint (Ito and Mester 2003 [1992]) requiring words to be structurally binary. As a result, prosodic words must be minimally disyllabic (hence no clipped words of monosyllabic size) and maximally bipodal (hence no clipped words larger than FF). Binarity requirements arguably hold at all other levels of the prosodic hierarchy (29). They are perhaps most well-established for foot structure (FtBin, Prince 1980), and are increasingly brought to bear on higher levels (φ-Bin, Kubozono 1988; Selkirk 2000). As is the case for Word Binarity for clipped words, precise formulations call for separate maximal and minimal versions (see Mester 1994, Selkirk 2000, and Ito and Mester 2007).

Word formation with a prosodic morphological target can thus reveal the phonologically optimal size and structure of prosodic words in a language, including alignment of prosodic word edges with foot edges, and binarity at the word level. For analytic details and further theoretical motivation, readers are referred to the work cited, as well later work where other generalizations emerged, such as Labrune's (2002) accent cut generalization (whereby loanwords are found to be truncated up to the accent), and pseudo-compound structuring documented by Sato (2002), Kubozono (2002), and others (where the second member of the (pseudo-)compound is treated as the truncated portion).

6 Summary and concluding remarks

This chapter started out by seeking an understanding of the phonological properties of affixal word formation (section 2), developing a phonological cross-classification of affixes, as well as an understanding of the types of phonological alternations observed (section 3), such as postnasal voicing, voicing assimilation, accent shifts, and phonologically motivated allomorphy. After outlining the phonological typology of compound structures and their prosodic implications (section 4), we turned to templatic word formations (section 5), where the shape and size of certain words (nicknames, clippings, reduplication, language games) are determined directly by phonological (prosodic) templates.

The coverage in terms of types of word formation has in no way been exhaustive. We have here focused our attention on those aspects that may not be dealt with in detail in other chapters of this volume, and that bear on cross-linguistic questions and typological issues as well as general theoretical ramifications and consequences. Many of these case studies of Japanese have served as major cornerstones in the development of general phonological theory, and others, we believe, are ripe for important future exploration.

Acknowledgements

For very helpful comments and suggestions on an earlier version, we are grateful to Shigeto Kawahara, Haruo Kubozono, and an anonymous reviewer.

References

Bach, Emmon and Robert T. Harms. 1972. How do languages get crazy rules? In Robert Stockwell and Ronald Macaulay (eds.), *Linguistic change and generative theory: Essays from the UCLA conference on historical linguistics in the perspective of transformational theory*, 1–21. Bloomington, Indiana: Indiana University Press.

Bat-El, Outi. 1996. Selecting the best of the worst: The grammar of Hebrew blends. *Phonology* 13(3). 283–328.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. San Diego: University of California dissertation.

Benua, Laura. 1995. Identity effects in morphological truncation. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers in Linguistics 18), 77–136. Amherst, MA: GLSA.

Bhat, D.N.S. 1978. A general study of palatalization. In Joseph H. Greenberg (ed.), *Universals of human language, vol. 2: Phonology*, 47–92. Stanford, CA: Stanford University Press.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bloch, Bernard. 1946a. Studies in colloquial Japanese I: Inflection. *Journal of the American Oriental Society* 66. 97–109.

Bloch, Bernard. 1946b. Studies in colloquial Japanese III: Derivation of inflected words. *Journal of the American Oriental Society* 66. 304–315.

Caballero, Gabriela. 2006. "Templatic back-copying" in Guarijío abbreviated reduplication. *Morphology* 16. 273–289.

Davis, Stuart and Natsuko Tsujimura. 1991. An autosegmental account of Japanese verbal conjugation. *Journal of Japanese Linguistics* 13. 117–144.

de Chene, Brent. 2010. Description and explanation in inflectional morphophonology: Notes on the Japanese verb. In Hiroki Maezawa and Azusa Yokogoshi (eds.), *Proceedings of the 6th Workshop on Altaic Formal Linguistics* (MIT Working Papers in Linguistics 61), 123–138. Cambridge, MA: MIT.

Downing, Laura. 2000. Morphological and prosodic constraints on Kinande verbal reduplication. *Phonology* 17. 1–38.

Gouskova, Maria. 2007. The reduplicative template in Tonkawa. *Phonology* 24. 367–396.

Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas A. Sebeok (ed.), *Current trends in linguistics*, 401–458. The Hague: Mouton.

Hashimoto, Shinkichi. 1948. Kokugogaku gairon [Introduction to Japanese Linguistics]. In Shinkichi Hashimoto (ed.), *Hashimoto Shinkichi hakushi chosakushū* 1 [Collected works by Shinkichi Hashimoto 1]. Tokyo: Iwanami.

Hattori, Shiro. 1973. Akusento towa nani ka. Sono benbetsuteki tokuchō towa [What is the prosodeme? And what is its distinctive feature?]. *Gengo no Kagaku* 4. 46–53.

Hyman, Larry. 2001. The limits of phonetic determinism: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*, 141–185. New York: Academic Press.

Inkelas, Sharon. 1989. *Prosodic constituency in the lexicon*. Stanford, CA: Stanford University dissertation.

Inkelas, Sharon and Cheryl Zoll. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Irwin, Mark. 2012. Rendaku dampening and prefixes. *NINJAL Research Papers* 4. 27–36.

Ishihara, Shinichiro. this volume. Chapter 14: Syntax-phonology interface. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko. 1986. *Syllable theory in prosodic phonology*. Amherst, MA: University of Massachusetts dissertation.

Ito, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–259.

Ito, Junko. 1990. Prosodic minimality in Japanese. In Karen Deaton, Manuela Noske and Michael Ziolkowski (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.

Ito, Junko, Yoshihisa Kitagawa and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Ito, Junko and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Cambridge, MA: Blackwell.

Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko and Armin Mester. 2003. Systemic markedness and faithfulness. In Jon Cihlar, Amy Franklin, Dave Kaiser and Irene Kimbara (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 39*, 665–689.

Ito, Junko and Armin Mester. 2003 [1992]. Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.), *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*, 26–65. Tokyo: Kaitakusha. [Appeared first as *UC Santa Cruz Linguistic Research Center Working Paper* LRC-92-09, 1992.]

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure* (MIT Press Linguistic Inquiry Monograph Series 41). Cambridge, MA: MIT press.

Ito, Junko and Armin Mester. 2004. Morphological contrast and merger: *Ranuki* in Japanese. *Journal of Japanese Linguistics* 20. 1–18.

Ito, Junko and Armin Mester. 2007. Prosodic adjunction in Japanese compounds. In Yoichi Miyamoto and Masao Ochi (eds.), *Formal approaches to Japanese linguistics: Proceedings of FAJL 4*, 97–111. Cambridge, MA: Department of Linguistics and Philosophy, MIT.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Junko, Armin Mester and Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26. 571–614.

Kageyama, Taro. 1976–77. Incorporation and Sino-Japanese verbs. *Papers in Japanese Linguistics* 5. 117–155.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82. 536–574.

Kawahara, Shigeto. 2012. Review of Laurence Labrune 2012. The phonology of Japanese. *Phonology* 29. 540–548.

Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Basic Blackwell.

Kim, Hyunsoon. 2001. A phonetically based account of phonological stop assibilation. *Phonology* 18(1). 81–108.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In In-Seok Yang (ed.), Linguistics in the morning calm, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17. 351–67.

Kochetov, Alexei. 2011. Palatalization. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1666–1690. Malden, MA and Oxford: Wiley-Blackwell.

Kubozono, Haruo. 1988. *The organization of Japanese prosody*. Edinburgh: University of Edinburgh dissertation.

Kubozono, Haruo. 1990. Phonological constraints on blending in English as a case for the phonology-morphology interface. *Yearbook of Morphology* 3. 1–20.

Kubozono, Haruo. 1995. *Gokeisei to on'in kōzō* [Word formation and phonological structure]. Tokyo: Kurosio.

Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo. 2002. Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Journal of the Phonetic Society of Japan* 6. 79–97.

Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In Caroline Féry and Ruben van de Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kubozono, Haruo, Junko Ito and Armin Mester. 1997. On'inkōzō kara mita go to ku no kyōkai: Fukugō meishi akusento no bunseki [The phonological boundary between the word and the phrase: An analysis of Japanese compound noun accentuation]. In Spoken Language Working Group (ed.), *Bunpō to onsei* [Speech and Grammar], 147–166. Tokyo: Kurosio.

Kurisu, Kazutaka. 2000. Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9. 147–185.

Kuroda, Shige-Yuki. 1965. *Generative grammatical studies in the Japanese language*. Cambridge, MA: MIT dissertation.

Labrune, Laurence. 2002. The prosodic structure of simple abbreviated loanwords in Japanese: A constraint-based account. *Journal of the Phonetic Society of Japan* 6. 98–120.

Labrune, Laurence. 2012. Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29. 113–52.

Lappe, Sabine. 2003. Monosyllabicity in prosodic morphology: The case of truncated personal names in English. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology* 2002, 135–86. Dordrecht: Kluwer.

Lombardi, Linda. 1998. Evidence for MaxFeature constraints from Japanese. University of Maryland Ms. ROA-247.

Martin, Samuel E. 1952. *Morphophonemics of standard colloquial Japanese*. Baltimore: Linguistic Society of America.

Martin, Samuel. 1975. *A reference grammar of Japanese*. New Haven and London: Yale University Press. Reprinted in 2004. University of Hawaii Press.

Mascaró, Joan. 1996. External allomorphy as emergence of the unmarked. In Jacques Durand and Bernard Laks (eds.), *Current trends in phonology: Models and methods*, 473–483. University of Salford, Manchester: European Studies Research Institute.

McCarthy, John J. 1981. The role of the evaluation metric in the acquisition of morphology. In C.L. Baker and John J. McCarthy (eds.), *The logical problem of language acquisition*, 218–248. Cambridge, MA: MIT Press.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25. 271–319.

McCarthy, John J., Wendell Kimper and Kevin Mullin. 2012. Reduplication in Harmonic Serialism. *Morphology* 22. 173–232.

McCarthy, John J. and Alan S. Prince. 1986. *Prosodic Morphology*. University of Massachusetts at Amherst and Brandeis University Ms. Reissued in 1996 as "Prosodic Morphology 1986." Technical Report #32, Rutgers University Center for Cognitive Science.

McCarthy, John J. and Alan S. Prince. 1990a. Prosodic morphology and templatic morphology. In Mushira Eid and John J. McCarthy (eds.), *Perspectives on Arabic linguistics II: Papers from the Second Annual Symposium on Arabic Linguistics*, 1–54. Amsterdam: John Benjamins.

McCarthy, John J. and Alan S. Prince. 1990b. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

McCarthy, John J. and Alan S. Prince. 1993. Generalized alignment. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology*, 79–153.

McCarthy, John J. and Alan S. Prince. 1999. Faithfulness and identity in prosodic morphology. In Harry van der Hulst, René Kager and Wim Zonneveld (eds.), *The prosody-morphology interface*, 218–309. Cambridge, U.K.: Cambridge University Press.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Mester, Armin. 1990. Patterns of truncation. *Linguistic Inquiry* 21. 475–485.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.

Mester, Armin and Junko Ito. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Miyagawa, Shigeru. 2012. Blocking and causatives: Unexpected competition across derivations. In Matthew A. Tucker, Anie Thompson, Oliver Northrup and Ryan Bennett (eds.), *Proceedings of the Fifth Formal Approaches to Japanese Linguistics Conference* (MIT Working Papers in Linguistics 64), 177–196. Cambridge, MA: MIT.

Nasu, Akio. this volume. Chapter 6: Phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Nespor, Marina and Irene Vogel. 1983. Prosodic structure above the word. In Anne Cutler and D. Robert Ladd (eds.), *Prosody: Models and measurements*, 123–140. Berlin, Heidelberg, New York, Tokyo: Springer.

Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Paster, Mary. 2006. *Phonological conditions on affixation*. Berkeley: University of California dissertation.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The prosody-morphology interface*, 310–343. Cambridge, MA: Cambridge University Press.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Poser, William J. 1984a. Hypocoristic formation in Japanese. In Mark Cobler, Suzannah MacKaye and Michael T. Wescoat (eds.), *Proceedings of WCCFL 3*, 218–229. Stanford: Stanford Linguistic Association.

Poser, William J. 1984b. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.

Poser, William J. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105.

Prince, Alan S. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Prince, Alan S. and Paul Smolensky. 1993 [2004]. *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University and University of Colorado, Boulder Ms RuCCS-TR-2. (Blackwell, 2004).

Sanders, Nathan. 1999. Same-edge alignment with opposite-edge effects. Talk presented at WCCFL 18 (http://sanders.phonologist.org/Papers/sanders-wccfl18.pdf, 5/5/2013).

Sano, Shin-Ichiro. 2012. Deriving global paradigm optimization in Japanese voice from constraint interaction. In Matthew A. Tucker, Anie Thompson, Oliver Northrup and Ryan Bennett (eds.), *Proceedings of the Fifth Formal Approaches to Japanese Linguistics Conference* (MIT Working Papers in Linguistics 64), 231–240. Cambridge, MA: MIT.

Sato, Hirokazu. 2002. Gairaigo ni okeru onsetsu fukugō e no kubunka to akusento [Decomposition into syllable complexes and the accenting of Japanese borrowed words]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 6. 67–78.

Scheer, Tobias. In press. How diachronic is synchronic grammar? Crazy rules, regularity and naturalness. In Patrick Honeybone and Joseph Salmons (eds.), *The Oxford handbook of historical phonology*. Oxford: Oxford University Press.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (ed.), *Nordic prosody*, 111–140. Trondheim: TAPIR.

Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax*, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Selkirk, Elisabeth. 2000. The interaction of constraints on prosodic phrasing. In Merle Home (ed.), *Prosody, theory and experiment: Studies presented to Gösta Bruce*, 231–261. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John A. Goldsmith, Jason Riggle and Alan C. L. Yu (eds.), *The handbook of phonological theory*, 2nd edition, 435–485. Oxford: Blackwell.

Steriade, Donca. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms., MIT.

Suzuki, Shigeyuki. 1996. Chapter 3: Gakkō bunpō hihan [Critique of school grammar]. In Shigeyuki Suzuki (ed.), *Keitairon: Josetsu* [Introduction to morphology], 195–312. Tokyo: Mugi Shobō.

Suzuki, Shinkichi. 1972. *Nihongo bunpō: Keitairon* [Japanese grammar: Morphology]. Tokyo: Mugi Shobō.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Tateishi, Koichi. 1989. Theoretical implications of the Japanese musician's language. In E. Jane Fee and Katherine Hunt (eds.), *Proceedings of WCCFL 8*, 384–398. Stanford: Stanford Linguistic Association.

Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13. 209–235. Amherst: GLSA.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge, U.K.: Cambridge University Press.

Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Wolf, Matthew Adam. 2008. *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. Amherst, MA: University of Massachusetts dissertation.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80. 73–97.

Timothy J. Vance

10 Rendaku

1 The rendaku alternations in modern Tokyo Japanese

The term *rendaku* refers to a set of pervasive morphophonemic alternations in modern Japanese, and the phenomenon has become widely known to linguists around the world in recent decades. An affected morpheme has one allomorph beginning with a voiceless obstruent and another allomorph (the voiced or rendaku allomorph) that appears, at least sometimes, non-word-initially in a compound or prefix+base combination, as in *tori* 'bird' and *oya+dori* 'parent bird'. No over-arching generalization accounts for when rendaku voicing occurs and when it does not, and many individual vocabulary items can occur either with or without the voicing. However, the likelihood of rendaku voicing is influenced by a number of well-known factors, and the behavior of speakers on experimental tasks indicates that tendencies in the existing vocabulary affect their responses.

The examples in (1) show the full range of the rendaku alternations.

(1)	a.	hune 'boat'	kawa+bune 'river boat'	[ɸ]~[b]
	b.	hire 'fin'	hara+bire 'belly fin'	[ç]~[b]
	c.	hako 'box'	hasi+bako 'chopstick box'	[h]~[b]
	d.	tama 'ball'	me+dama 'eyeball'	[t]~[d]
	e.	ti 'blood'	hana+zi 'nosebleed'	[tʃ]~[dʒ]
	f.	tuka 'mound'	ari+zuka 'anthill'	[ts]~[(d)z]
	g.	sora 'sky'	hosi+zora 'starry sky'	[s]~[(d)z]
	h.	sima 'stripe'	yoko+zima 'horizontal stripe'	[ʃ]~[dʒ]
	i.	kami 'paper'	kabe+gami 'wallpaper'	[k]~[g]

Notice that [b] alternates not with [p] but with [ɸ] in (1a), with [ç] in (1b) and with [h] in (1c). The [ɸ]~[b], [ç]~[b], and [h]~[b] alternations are due to a well-known historical change: initial [ɸ], [ç], and [h] in native Japanese words are all descended from a single phoneme that was once pronounced [p] (see Takayama, this volume, and various chapters in the History Volume for more details). Notice also that [(d)z] alternates both with [ts] in (1f) and with [s] (1g), and that [dʒ] alternates both with [tʃ] (1e) and with [ʃ] (1h).[1] These pairings reflect mergers of voiced fricatives and affricates: Tokyo Japanese has lost earlier phonemic distinctions between [z] and [dz] and between [ʒ] and [dʒ] (Takayama, this volume). Because of all these changes,

[1] The alveopalatal consonants represented as [ʃ tʃ dʒ] in this volume are more accurately transcribed as [ɕ cɕ ɟʑ] (Vance 2008: 14, 82, 84).

the difference between the two allomorphs of an alternating morpheme is often more than just the presence or absence of voicing.[2]

It is not immediately obvious that all the alternations in (1) should be treated as instances of a single phenomenon. The same sort of problem arises in connection with the three voiceless/voiced fricative alternations in English nouns shown in (2).

(2) a. /f/~/v/ wolf wolves b. /θ/~/ð/ bath baths c. /s/~/z/ house houses

These three English pairings are all phonetically parallel, but many noun morphemes end in /f/ or /θ/ both in the singular and in the plural (e.g., *cliff/cliffs* and *lath/laths*), *house* (2c) is the only morpheme that shows the /s/~/z/ alternation, and no morpheme shows a parallel /ʃ/~/ʒ/ alternation. It is not at all certain that ordinary native speakers of English intuitively recognize the three fricative alternations as instances of a single more abstract phenomenon.

When it comes to rendaku, however, there is no real doubt that native speakers of Tokyo Japanese see all the alternations in (1) as instances of a single more general phenomenon, in spite of the phonetic complications. One likely reason is that the Japanese rendaku alternations are much more widespread than the English fricative alternations. The Japanese alternations appear in a very large number of morphemes, whereas the English alternations are confined to a small set of noun morphemes. At the same time, almost any preceding compound element or prefix provides an environment for the voiced allomorph (i.e., the allomorph showing rendaku) of an alternating Japanese morpheme. In the English case, the plural morpheme is the only environment for the allomorphs ending with a voiced fricative (assuming that noun–verb pairs like *belief* and *believe* are not instances of the same phenomenon).

The Japanese writing system provides what is undoubtedly an even more powerful reason for native speakers to see the rendaku alternations as a unitary phenomenon: all the alternations are represented in exactly parallel fashion in modern *kana* spelling, which was first adopted in 1946 (Yoshida and Inokuchi 1962: 667–684) and reaffirmed 40 years later with only minor modifications (Bunkachō 1986). The *kana* voicing diacritic (*dakuten*) represents more than just the addition of voicing in some cases, and the relationships between *kana* letters with and without *dakuten* mirror the alternations shown in (1) above. The examples in (3) illustrate.

(3) た ta さ sa か ka は ha
だ da ざ za が ga ば ba

2 The phonemic transcription and romanization systems adopted in this volume do not reflect these changes and may be inappropriate for modern Tokyo Japanese (see Pintér, this volume, for details). For speakers who have syllable-initial [ŋ] (Hibiya 1999; Vance 2008: 214–222), the rendaku partner of [k] is (or can be) [ŋ], which is one more deviation from simple presence versus absence of voicing.

Because of the mergers of voiced fricatives and affricates mentioned above, each of the syllables *zu zi zya zyo zyu* has two possible spellings. In most cases, the diacritic is added to the letters for *su si sya syo syu* to write *zu zi zya zyo zyu*, as in (4a). But when the voiceless allomorph (i.e., the allomorph without rendaku) of a morpheme begins with one of *tu ti tya tyo tyu*, the practice is to represent its voiced allomorph by just adding the diacritic to write *zu zi zya zyo zyu*, as in (4b).

(4)	a.	す su	し si	しゃ sya (spelled si_{ya})	etc.
		ず zu	じ zi	じゃ zya (spelled zi_{ya})	etc.
	b.	つ tu	ち ti	ちゃ tya (spelled ti_{ya})	etc.
		づ zu	ぢ zi	ぢゃ zya (spelled zi_{ya})	etc.

For example, *yoko+zima* 'horizontal stripe' in (1h) is spelled よこじま, with じ for *zi* [dʒi], because *sima* 'stripe' is spelled しま, with し for *si* [ʃi]. On the other hand, *hana+zi* 'nosebleed' in (1e) is spelled はなぢ, with ぢ for *zi* [dʒi], because *ti* 'blood' is spelled with ち for *ti* [tʃi].

The remainder of this chapter provides an up-to-date survey of research on rendaku. Section 2 sketches a scenario for how the rendaku alternations developed historically. Section 3 examines phonological factors that affect how likely rendaku is to occur, and section 4 assesses the extent to which the rendaku alternations have spread from the native vocabulary into other strata. Section 5 looks at semantic and morphological factors that seem to inhibit or promote rendaku, and section 6 argues that even when all known factors are taken into account, rendaku is, to a significant extent, unpredictable.

2 The historical development of rendaku

There is a plausible story about the historical origin of rendaku that involves prenasalization. It is generally accepted that voiced obstruents in Old (8th-century) Japanese (OJ) were prenasalized: [^{m}b ^{n}d ^{n}(d)z ^{ŋ}g] (Takayama, this volume). Prenasalization disappeared long ago in Tokyo and Kyoto Japanese, but an early 17th-century description by the Portuguese missionary João Rodrigues makes it clear that prenasalization was still present to some extent in Kyoto at that time (Hashimoto 1932; Irwin and Narrog 2012: 250). Prenasalization is still preserved even today in some dialects, most famously those of the Tōhoku (northeastern Honshū) region (Shibatani 1990: 204–205).

Example (5) and others like it (Hashimoto 1932: 5–6; Hamada 1952: 23; Frellesvig 2010: 42–43) show how a well-known type of historical change makes sense if voiced obstruents were prenasalized. (EMJ is Early Middle Japanese [800–1200], and MT is modern Tokyo Japanese.)

(5) $^{\text{EMJ}}$sumi+suri > $^{\text{MT}}$suzuri ‘inkstone’ (cf. $^{\text{MT}}$sumi ‘india ink’, $^{\text{MT}}$suri ‘rubbing)
[sumisuɾi] > [sũmsuɾi] > [sũmzuɾi] > [sũn(d)zuɾi] (= $^{\text{EMJ}}$*suzuri*)

The etymology in (5) is uncontroversial, and the earliest attestations in Nihon Kokugo Dai-jiten Dainihan Henshū Iin-kai (2000–02) (NKD hereafter) are 934 for $^{\text{EMJ}}$*sumi+suri* and late 10th century for $^{\text{EMJ}}$*suzuri*. The two forms were presumably in competition for a time, but the contracted form eventually prevailed. As the second line in (5) shows, the change from $^{\text{EMJ}}$/mis/ to $^{\text{EMJ}}$/z/ is easy to understand if $^{\text{EMJ}}$/z/ was prenasalized. The first step in the process was the loss of the vowel between the nasal [m] and the voiceless obstruent [s] – an unremarkable rapid-speech reduction that resulted in salient nasalization on the vowel that was now followed by a coda nasal consonant. The second step was the spread of voicing into the onset following the nasal consonant. The third step was the assimilation of the nasal to the place of articulation of the following onset consonant. At this point, listeners could reinterpret the phonetic sequence [ũn(d)z] as the realization of “underlying” $^{\text{EMJ}}$/uz/, since $^{\text{EMJ}}$/z/ and other voiced obstruents were realized with prenasalization. It is uncertain how many separate stages were really involved in this process and what order they occurred in, but something like the last line in (5) is a believable scenario. Since phonemic coda nasals were still not permissible (at least in the colloquial vocabulary) at this time (Takayama, this volume), it was not possible to reinterpret [ũn(d)z] as something like $^{\text{MT}}$/uNz/.[3] Modern Tokyo [sɯ(d)zɯɾi] (= $^{\text{MT}}$/suzuri/) reflects the later loss of prenasalization. (The more rounded high back vowel generally assumed for EMJ is unimportant here.) This correspondence between $^{\text{EMJ}}$/mis/ and $^{\text{MT}}$/z/ is just one instance of the general pattern: an earlier sequence of a nasal consonant (N) followed by a vowel (V) followed by a voiceless obstruent (T) corresponds to a modern Tokyo voiced obstruent (D) with the same place of articulation as the original voiceless obstruent: NVT > D.

The proposed explanation for the origin of rendaku depends on the reasonable assumption that the consonants corresponding to modern Tokyo voiced obstruents were prenasalized in prehistoric (pre-Old) Japanese as well. It also depends on the uncontroversial assumption that pre-OJ (like OJ) did not permit coda nasals (or any

3 Martin (1987: 125) says that coda nasals were well established in the colloquial language by 1200, so it eventually did become possible to reinterpret a phonetic sequence of a nasalized vowel (Ṽ) followed by a nasal consonant (N) followed by a voiced obstruent (D) as involving an underlying (i.e., phonemic) nasal consonant: [ṼND] ← /VND/. Consequently, later instances of the same kind of phonetic reduction led in many cases to what is traditionally called *onbin* ‘euphonic change’ (Takayama, this volume), as in $^{\text{EMJ}}$/yom-ite/ > $^{\text{MT}}$/yoN-de/ ‘reading’. Frellesvig (2010: 185–199) provides a concise account of *onbin* phenomena, and he notes the relevance of the development of phonemic coda nasals: “Some examples in OJ of syllable loss seem to involve the same kind of phonetic reduction as was involved in *onbin* ... The main difference between the developments ... is that no moraic phoneme arose in the examples from OJ ... This suggests that the phonetics which in the transition between OJ and EMJ gave rise to *onbin* already were a feature of OJ” (Frellesvig 2010: 198).

other coda consonants for that matter). As an illustration, consider the Modern Tokyo compound *asa+giri* ‘morning fog’. The OJ ancestor of this word is attested, and it had rendaku: $^{\mathrm{OJ}}$*asa+gwiri*.[4] Compare $^{\mathrm{MT}}$*asa* ‘morning’, corresponding to $^{\mathrm{OJ}}$*asa*, and $^{\mathrm{MT}}$*kiri* ‘fog’, corresponding to $^{\mathrm{OJ}}$*kwiri*. The voiced obstruent in $^{\mathrm{OJ}}$*asa+gwiri* was realized with prenasalization: $^{\mathrm{OJ}}$/asa+[$^{\mathrm{n}}$g]wiri/. Assuming prenasalization in pre-OJ, and given the natural development NVT > D, it makes sense to infer that $^{\mathrm{OJ}}$*asa+gwiri* developed from an ancestor of the form $^{\mathrm{pre\text{-}OJ}}$/asa/+NV+/kwiri/. The obvious candidate for the NV syllable here is the ancestor of the OJ genitive particle $^{\mathrm{OJ}}$*no* (cf. $^{\mathrm{MT}}$*no*), as in (6) (Murayama 1954: 107; Unger 1975: 8–9; Vance 1982: 335–338; Frellesvig 2010: 40–43).

(6) ($^{\mathrm{pre\text{-}OJ}}$/asa+no+kwiri/) > $^{\mathrm{OJ}}$/asa+gwiri/[asã$^{\mathrm{n}}$gwiri] > $^{\mathrm{MT}}$/asa+giri/
/nok/ > /g/
NVT > D

The prehistoric form in parentheses on the left in the top line in (6) is, of course, hypothetical.[5]

On the other hand, there is no reason to assume that every OJ noun+noun compound noun developed from an ancestor of the form noun+$^{\mathrm{pre\text{-}OJ}}$/no/+noun. Modern Tokyo Japanese has frozen noun+*no*+noun phrases like *te+no+hira* ‘palm of the hand’ (containing *te* ‘hand’ and *hira* ‘flat’) alongside simple noun+noun compounds like *te+kubi* ‘wrist’ (containing *kubi* ‘neck’). The situation in prehistoric Japanese was probably much the same. Consider the three attested OJ examples in (7). The forms in parentheses on the left are hypothetical prehistoric forms.[6]

4 OJ examples are romanized using the system adopted by Frellesvig and Whitman (2008: 2–5).
5 Kupchik (2012) argues that the Eastern OJ poems recorded in the *Man’yōshū* provide evidence that this kind of NVT to D contraction was an optional synchronic process in the eastern dialects and that poets could exploit it to shorten a line and make it fit the meter.
6 The examples in (7) are all attested in phonograms in extant copies of 8th-century texts. The *kanji* used phonographically in OJ texts are called *man’yōgana* (Takayama, this volume). Many OJ words are attested only in *kanji* used logographically, which means there is no direct evidence for how they were pronounced. Of course, since only later copies of the principal OJ texts exist, the phonogram representations were susceptible to copying errors and misguided “corrections” by later scribes. The modern Tokyo descendant of (7a) is *kaede* ‘maple’, a reduced form of *kaeru+de*, which is now obsolete but had developed rendaku. The first element in (7b), $^{\mathrm{OJ}}$*tama* ‘jewel’, was frequently added to nouns as a kind of honorific. This compound has not survived into modern Tokyo Japanese. Phrasal $^{\mathrm{OJ}}$*kwo+no+te* in (7c) is attested only as the first part of the tree name $^{\mathrm{OJ}}$*kwo+no+te+kasi+pa* ‘oriental arbor-vitae’. This definition is the species denoted by the modern Tokyo descendant of this tree name: $^{\mathrm{MT}}$*ko+no+te+gasiwa*, with rendaku in the last element (cf. $^{\mathrm{MT}}$*kasiwa* ‘oak’). Although $^{\mathrm{MT}}$*kasiwa* is etymologically a compound, as indicated by the morpheme boundary in corresponding $^{\mathrm{OJ}}$*kasi+pa* (cf. $^{\mathrm{OJ}}$*kasi* ‘oak’, $^{\mathrm{OJ}}$*pa* ‘leaf’), ordinary present-day speakers think of it as monomorphemic.

(7) a. ($^{\text{pre-OJ}}$/kaperu+te/ >) $^{\text{OJ}}$kaperu+te 'maple' (lit. 'frog hand')
b. ($^{\text{pre-OJ}}$/tama+no+te/ >) $^{\text{OJ}}$tama+de 'jewel-like hand'
c. ($^{\text{pre-OJ}}$/kwo+no+te/ >) $^{\text{OJ}}$kwo+no+te 'child's hand'
(in $^{\text{OJ}}$kwo+no+te+kasi+pa)

The idea is that some prehistoric noun+$^{\text{pre-OJ}}$/no/+noun combinations remained phrases in outward form (like $^{\text{OJ}}$*kwo+no+te*), while others contracted and developed into compounds with rendaku (as in $^{\text{OJ}}$*tama+de*). Meanwhile, combinations formed by simple juxtaposition remained compounds without rendaku (like $^{\text{OJ}}$*kaperu+te*). If the proposed account of the origin of rendaku is correct, these examples show why we would expect the phenomenon to be as irregular as it was in OJ. Some OJ compounds (those with rendaku) had developed from phrases, while others (those without rendaku) had been formed by simple juxtaposition.

When it comes to reduplicated words (see section 4.2), the ancestor of the modern dative particle *ni* seems like a more likely candidate than a genitive particle for an NV syllable that could have contracted and left rendaku as a residue (Lyman 1894: 13; Martin 1987: 104). As Frellesvig (2010: 40–41) points out, however, there are examples of rendaku in OJ that do not seem to be derivable from any earlier phrase with an NV syllable between the elements, and he draws the reasonable conclusion that "*rendaku* already in OJ was established as a morphophonemic process" that could trigger irregular analogical extensions. The most notorious example is $^{\text{OJ}}$*ama+no+gapa* 'Milky Way' (cf. $^{\text{OJ}}$*ama~ame* 'heaven', $^{\text{OJ}}$*kapa* 'river'; modern Tokyo *ama+no+gawa*). This frozen phrase had rendaku in OJ even though the genitive particle had not contracted. The irregularities in OJ could have been leveled out in the subsequent 1,300 years, but this has not happened. Modern Tokyo Japanese is little different overall, although many individual vocabulary items have gained or lost rendaku over the centuries.

3 Phonological factors affecting rendaku

3.1 Lyman's Law

A non-initial voiced obstruent in a morph inhibits rendaku. For example, compare *umi+kaze* 'sea breeze' and *umi+game* 'sea turtle'. The independent words *kaze* 'wind' and *kame* 'turtle' both begin with voiceless /k/, but *kaze* contains /z/, which is realized as a voiced obstruent ([dz] or [z]).[7] The idea is that the /z/ in *kaze* prevents rendaku and rules out the form **umi+gaze*. This apparent constraint on rendaku is usually called Lyman's Law, in honor of the American geologist who provided

7 On the distribution of the allophones of /z/, see Maekawa (2010).

the first thorough account of it (Lyman 1894).[8] The history of Lyman's research on Japanese is now well understood (Ogura 1910; Vance 2007a, 2012), and although there are terse references to the constraint in earlier Japanese sources (Miyake 1932; Suzuki 2007: 232), there is little doubt that Lyman discovered it independently.

It has been suggested that a voiced obstruent has this inhibiting effect only if it is in the mora immediately following the potential rendaku site and not if it is in a later mora (Okumura 1955; Nakagawa 1966: 302; Sakurai 1966: 41; Maeda 1977). According to this restricted version of Lyman's Law, rendaku would not affect morphemes like *suzume* 'sparrow' and *kuzira* 'whale', but it might occur in morphemes like *tokage* 'lizard' and *hituzi* 'sheep'. The existing vocabulary does not suggest that Lyman's statement needs to be weakened by incorporating a distance effect. Iwanami Shoten Jiten Henshūbu (1992), a large reverse-lookup dictionary that includes many obscure and obsolete words among its entries, lists 26 words ending in *suzume*, 19 ending in *kuzira* 'whale', nine ending in *tokage*, and four ending in *hituzi*. There are no entries ending in the hypothetical voiced allomorphs: **zuzume*, **guzira*, **dokage*, **bituzi*. The examples in (8) are representative.

(8) umi+suzume 'murrelet' (cf. umi 'sea')
ha+kuzira 'toothed whale' (cf. ha 'tooth')
doku+tokage 'poisonous lizard' (cf. doku 'poison')
ko+hituzi 'lamb' (cf. ko 'child')

These examples suggest that it is just the presence of a voiced obstruent in a morph and not its location that matters. That is, Lyman's Law seems to be a constraint that prevents rendaku whenever the voiced allomorph of a morpheme would contain any voiced obstruent other than the morph-initial one, and this is how Martin (1952: 48) describes it. As explained later in this section, there are theoretical accounts of Lyman's Law that see it as an automatic consequence of a constraint limiting voiced obstruents to one per morph. These accounts depend crucially on the version of Lyman's Law assumed here, i.e., that a non-initial voiced obstruent anywhere in a morph inhibits rendaku.

Despite the lack of evidence for a distance effect in the existing vocabulary, a nonce-word survey carried out in the late 1970s provided weak support for such an effect as a factor in the behavior of some speakers (Vance 1979: 100–106, 1980b: 258–259). The participants were asked to choose between a pronunciation with rendaku

8 The statement of Lyman's Law given here differs in one significant way from what Lyman (1894: 2) actually wrote: he included /p/ on his list of consonants that inhibit rendaku. Needless to say, modern Tokyo /p/ is not realized as a voiced obstruent, and Ogura (1910: 11) suggested that Lyman must have been misled by traditional Japanese terminology, in which a mora beginning with a voiced obstruent is called a *dakuon* and a mora beginning with /p/ is called a *han-dakuon* 'half *dakuon*'.

and a pronunciation without rendaku for examples like *kawa+sabari* vs. *kawa+zabari*, *kawa+sotogi* vs. *kawa+zotogi*, and *kawa+sawasobi* vs. *kawa+zawasobi*. Some of the participants did show a statistically significant distance effect, but the effect size was small, and the test items were not designed to avoid possible confounds. Ihara and Murata (2006: 21–22) report that larger-scale studies done in 1984 and 2005 were able to replicate the effect, but a study by Kawahara (2012), using a different methodology, found no effect. The apparent distance effect in the earlier studies was probably due to some other variable that was uncontrolled.

Compounds that consist of more than two elements raise an interesting question about Lyman's Law. Each of the two three-element compounds in (9) contains a morpheme meaning 'fire' as its second element, but the semantic constituent structure of two compounds differs. (These compounds are obscure words, but both are listed in authoritative dictionaries, and they provide a nice illustration of a general pattern.)

(9) a. {{kagari+bi}+bana} 'cyclamen'
cf. kagari 'iron fire basket', hi 'fire', kagari+bi 'bonfire'
hana 'flower'

b. {hako+{hi+bati}} 'boxed brazier'
cf. hako 'box'
hi 'fire', hati 'bowl', hi+bati 'brazier'

Examples like these suggest that Lyman's Law applies to each layer of compounding. In (9a), the inner (lower) layer is *kagari+bi*, and since *hi~bi* 'fire' contains no non-initial voiced obstruent, *kagari+bi* does not violate Lyman's Law. At the outer (higher) layer, *hana~bana* 'flower' is the second element, and since it contains no non-initial voiced obstruent, there is no Lyman's Law violation in {{*kagari+bi*}+*bana*}. In (9b), the inner layer is *hi+bati*, and since *hati~bati* 'bowl' contains no non-initial voiced obstruent, *hi+bati* does not violate Lyman's Law. But *hi+bati* is the second element at the outer layer, and it does contain a non-initial voiced obstruent, so *{*hako*+{*bi+bati*}} would be a violation, assuming that Lyman's Law applies to each layer of compounding. There is, however, an alternative explanation, since *{*hako*+{*bi+bati*}} also violates what is known as the right-branch condition (see section 5.2).

Returning now to two-element compounds, the modern Tokyo vocabulary contains a very small number of well-known exceptions to Lyman's Law. First, when a compound contains *hasigo~basigo* 'ladder' as its second element, the voiced allomorph *basigo* ordinarily appears, as in *nawa+basigo* 'rope ladder' (cf. *nawa* 'rope'). Although the etymology of this element meaning 'ladder' is uncertain (Martin 1987: 115), the usual *kanji* spelling (梯子) invites a literate native speaker to analyze it as a compound containing the voiced allomorph of *ko~go* 'child' as its second element. But even if *hasigo~basigo* is itself a compound, it is clearly the inner layer

in *nawa+basigo*: {*nawa*+{*basi+go*}}. At the outer layer, the second of the two constituents is *basi+go*, which contains a non-initial voiced obstruent. As noted just above, the general pattern in a three-element compound of this form is for the second element not to show rendaku.[9]

The voiced obstruent /g/ is in the third mora of *hasigo~basigo* 'ladder', but other exceptions to Lyman's Law have a voiced obstruent in the second mora of the element that shows rendaku. A traditional name for a third son is *saburoo*, but many third sons have names that begin with an additional element, and many of these longer names show rendaku (Kindaichi 1976: 5), as in *ken+zaburoo*. The name *saburoo* itself is easily analyzable into two morphs: *sabu+roo*. The second realizes a morpheme that appears in other masculine names that reflect birth order: *iti+roo* (for a first son), *zi+roo* (for a second son), and so on. If we treat *ken+zaburoo* as {*ken*+{*zabu+roo*}}, it contains the non-initial voiced obstruent /b/ in the second of the two constituents at the outer layer of compounding. Once again, the general pattern in a three-element compound of this form is for the second element not to show rendaku.

One other exception to Lyman's Law is the slang verb *hun+zibar-u* 'tie up roughly' (Kindaichi 1976: 5). According to one possible analysis, this word consists of an unproductive prefix *hun-* 'roughly' (etymologically a reduced form of the verb element *humi*; cf. *hum-u* 'step on'), an allomorph of the same verb root that appears in *sibar-u* 'tie up', and an inflectional ending (nonpast affirmative *-u* in the citation form). This analysis gives us the stem *hun+zibar* for the slang verb, and the non-initial voiced obstruent in the second morph makes the morph-initial voiced obstruent an obvious violation of Lyman's Law.[10] Other reported exceptions to Lyman's Law (Kindaichi 1976: 5; Martin 1987: 115; Vance 1987: 137; Suzuki 2005) are either dubious or obsolete.

Ramsey and Unger (1972: 287–289) say that rendaku did not occur in OJ if either the first or the second element in a two-element compound contained a voiced obstruent. Unger (1975: 9) calls this the "strong version" of Lyman's Law, and there

9 It is likely that all the three-syllable second elements in (8) above had morphologically complex ancestors, but even on the implausible assumption that that they are synchronic compounds, Lyman's Law, applied at each layer of compounding, predicts that they should resist rendaku. Compare the etymological compounds *koto+ba* 'language' (cf. *koto* 'word', *ha* 'leaf'), which is probably about as hard for a modern Tokyo speaker to analyze as *hasi+go*, and *tama+go* 'egg' (cf. *tama* 'ball, *ko* 'child'), which is more transparent. As Lyman's Law predicts, these two items do not show rendaku as constituents in longer compounds, e.g., {*kuti*+{*koto+ba*}} 'spoken language' (cf. *kuti* 'mouth') and {*nama*+{*tama+go*}} 'raw egg' (cf. *nama* 'raw').

10 The morphemic divisions of inflectional forms cited here follow the widely adopted analysis of Bloch (1946). These divisions are used just for convenience and are not intended to imply an endorsement of the analysis behind them. Verb forms in particular raise problems for morphemic analysis that need not be resolved here (Vance 1987: 175–208, 1991; Klafehn 2003). Nothing of any consequence here turns on the division between stem and ending in *hun+zibar-u*.

are no persuasive OJ examples that violate it (Vance 2005b: 32–33). The strong version of Lyman's Law clearly does not apply to modern Tokyo Japanese, since it is easy to find counterexamples like *sode+guti* 'cuff' (cf. *sode* 'sleeve', *kuti* 'mouth'), *kagi+zume* 'hooked claw' (cf. *kagi* 'hook', *tume* 'claw'), and *tabi+bito* 'wayfarer' (cf. *tabi* 'journey', *hito* 'person').

Lyman's Law has attracted the attention of theoretical phonologists, and it has become very popular in recent years to interpret it as a manifestation of a much more general constraint called the obligatory contour principle (OCP). Leben originally proposed the OCP in 1973 as a prohibition against sequences of identical tones in underlying representations (Leben 2011: 326), and it was incorporated into early autosegmental phonology as a restriction on the tonal tier (Goldsmith 1990: 309–318). Once the idea of putting non-tonal features on separate tiers caught on, allowing such features to be treated as effectively suprasegmental, it was possible to formulate a ban on multiple occurrences of voicing by setting up a voicing tier and invoking the OCP to rule out adjacent identical specifications (Odden 2011: 22).

Rendaku itself can be interpreted as the introduction of a voicing specification whenever a compound is formed. Ito and Mester (1986: 56–57) treat rendaku as the realization of a linking morpheme that joins the two members of a compound. Phonologically, this morpheme is a floating voicing specification that Ito and Mester see as the synchronic residue of the NV syllable that contracted and produced the original instances of rendaku in prehistoric Japanese (see section 2).[11] Of course, the feature involved has to be something more abstract than just phonetic voicing; it has to convert a voiceless obstruent into the voiced obstruent that rendaku pairs it with, as in (1) above. Also, this floating specification can only dock onto an obstruent at the beginning of the second element of a compound. Otherwise, it could produce forms like **izi+muro* instead of *isi+muro* 'stone hut' (cf. *isi* 'stone', *muro* 'dwelling') or **yama+nego* instead of *yama+neko* 'wildcat' (cf. *yama* 'mountain', *neko* 'cat'). Kuroda (2002) treats this voicing feature as linked underlyingly to the initial obstruent of a morpheme that has a rendaku allomorph, and this link has to be severed when such a morpheme appears word-initially. Ito and Mester (1986: 57–60) adopt a rightward voicing spread rule that cannot affect or skip over a vowel or sonorant without violating the general constraints on rule application that they assume.

Once this basic machinery is in place, some additional assumptions are necessary to make the OCP approach work for Lyman's Law. First, the voicing specification has to be limited to obstruents, that is, the only segment type where the presence versus absence of voicing is distinctive in Japanese. If the non-distinctive voicing of vowels and sonorant consonants were specified too, any morph with two or more voiced segments in a row (like *hiza* 'knee' or *sune* 'shin') would violate the OCP.

11 Hirano (1974: 35–38) suggests that an element like this linking morpheme can be inferred for some earlier stage of Japanese via internal reconstruction. He calls this element a "ligateme" because of its similarity to what grammars of Tagalog call a ligature.

Second, voicing has to be a monovalent feature, not a traditional binary feature that specifies voiced obstruents as [+vce] and voiceless obstruents as [–vce] (Mester and Ito 1989: 277–279). Otherwise, rendaku in a case like **ko+bituzi* (instead of the actual form *ko+hituzi* 'lamb') would not be ruled out by the OCP, since the [+vce] specifications associated with *b* and *z* would be separated by the [–vce] specification associated with *t*.

Under these assumptions, Lyman's Law works as shown in (10). These simplified autosegmental representations show a separate voicing tier but all other segmental information consolidated into a single tier. Rendaku is allowed in *ao+zame* 'blue shark' but not in *ao+sagi* 'blue heron'.

(10) a. (rendaku)
[vce]
|
a o + s a m e
'blue' + 'shark'

b. (rendaku)
[vce] [vce]
|
a o + s a k i
'blue' + 'heron'

Since there are two adjacent [vce] specifications in (10b), the OCP prevents the unlinked one from linking and ensures that the compound meaning 'blue heron' does not surface as **ao+zagi*. The domain of the OCP has to be limited, of course, so that it only applies to voicing specifications that would otherwise end up linked to segments in the same morph. Without this limitation, any word containing more than one voiced obstruent would be a violation, including *kage+guti* 'backbiting' (cf. *kage* 'shade', *kuti* 'mouth') and even *nozoki+mado* 'observation window' (cf. *nozok-u* 'peek', *mado* 'window').

In an optimality theory (OT) approach, Lyman's Law can be implemented by taking advantage of a generally accepted markedness constraint prohibiting voiced obstruents: VOP (voiced obstruent prohibition), also known as NO-D (Ito and Mester 2003: 26). The idea that voiced obstruents are more marked than voiceless obstruents is not controversial (Ohala 1983: 194–202), but in a language that has distinctively voiced obstruents, like modern Tokyo Japanese, this constraint has to be ranked lower than IDENT-IO(voice), a faithfulness constraint that requires a voiced obstruent in the input to be preserved in the output (Ito and Mester 2003: 36). One proposal that has been advanced for implementing Lyman's Law (and other OCP effects) is known as self-conjunction (Alderete 1997; Kager 1999: 397–399). Ito and Mester (1998: 4) explain as follows. "The central idea is that *there is no Obligatory Contour Principle* per se: Universal Grammar is not concerned about adjacent identicals qua identicals. Rather, OCP-effects arise when markedness constraints are violated more than once." VOP^2, the self-conjunction of VOP, penalizes an output that contains two violations of VOP, and it can be treated as a separate constraint and ranked independently (Ito and Mester 2003: 36–38). Since the domain of Lyman's Law construed as an OCP effect is the morph, the effect of VOP^2 has to be limited to the morph, and

Ito and Mester (1998: 32, 2003: 105–108) have proposed two different ways of accomplishing this. Then, by ranking IDENT-IO(voice) below morph-domain VOP² but above VOP, a candidate containing a morph with two voiced obstruents, like **ao+zagi* in (10b) can be ruled out, while a candidate containing a morph with a single voiced obstruent that is present in the input (like *sagi* 'heron') can emerge as a winner (Ito and Mester 2003: 37). Not surprisingly, the notion of self-conjunction has not been greeted with universal enthusiasm even among OT advocates, and Ito and Mester (2003: 59–61) offer an interesting discussion of some of the problems involved.

Rendaku itself can be reformulated in OT by retaining the idea that it is the realization of a linking morpheme and invoking a constraint requiring every morpheme to have some kind of exponent – a constraint that is often called REALIZE-MORPHEME (Kurisu 2001: 37–56; Ito and Mester 2003: 81–97). REALIZE-MORPHEME must be ranked lower than the morph-domain VOP² that enforces Lyman's Law but high enough to ensure that rendaku appears when it does not result in a Lyman's Law violation. Of course, as mentioned above, there has to be some way to ensure that the realization can only appear in an obstruent at the beginning of the immediately following morph. Ito and Mester (2003: 83–84) assume that the rendaku morpheme acts like a prefix to the following element, forming a constituent with it, and they attribute the restriction on where it can be realized to other constraints that are necessary anyway (Ito and Mester 2003: 88).

Theoretical phonologists who know modern Tokyo Japanese are well aware that rendaku is pervasively irregular, as noted above in section 1 and explained in more detail below in section 3.2. Ito and Mester (2003: 83–85) take the sensible position that this irregularity is for the most part a matter of the presence or absence of the linking morpheme that rendaku realizes in their account. In other words, they attribute the irregularity to morphology, allowing the phonological analysis they propose to apply largely unimpeded. This strategy does not offer any explanation for the persistent irregularities that characterize rendaku, but it shifts the responsibility for explaining them (to the extent that any sort of explanation is possible) outside of phonology.

3.2 Rosen's Rule

Rendaku is largely unpredictable because of two basic types of irregularity. First, certain elements are rendaku-immune, i.e., they never show rendaku, even when no inhibiting factor is involved (see, e.g., Irwin 2009: 192–193). Second, many other elements behave inconsistently; they sometimes show rendaku but often do not, even when no inhibiting factor is involved.

Some native Japanese noun morphemes that are immune to rendaku are listed in (11). The reason for restricting the list to native nouns is that items in this class are in general the most likely to show rendaku (see section 4 and section 5.3).

(11) kemuri 'smoke', as in suna+kemuri 'clouds of sand' (cf. suna 'sand')
tuyu 'dew', as in asa+tuyu 'morning dew' (cf. asa 'morning')
himo 'string', as in kutu+himo 'shoelace' (cf. kutu 'shoe')

None of these immune morphemes contains a medial voiced obstruent, so rendaku would not violate Lyman's Law.[12]

The examples in (12) show native Japanese noun morphemes that behave inconsistently with respect to rendaku.

(12) a. ki 'tree; wood'
tumi+ki '(toy) wooden blocks' (cf. tum-u 'stack')
yose+gi 'wooden mosaic' (cf. yose-ru 'bring together')

b. sima 'island'
uki+sima 'floating island' (cf. uk-u 'float')
hanare+zima 'solitary island' (cf. hanare-ru 'be apart')

c. te 'hand'
hidari+te 'left hand' (cf. hidari 'left')
usiro+de 'hands behind one's back' (cf. usiro 'rear')

d. tama 'ball'
mizu+tama 'water droplet' (cf. mizu 'water')
yu+dama 'bubbles in boiling water' (cf. yu 'hot water')

e. hi 'sun'
yuu+hi 'evening sun' (cf. yuu 'evening')
nisi+bi 'westering sun' (cf. nisi 'west')

Some inconsistent morphemes show rendaku in a large majority of the relevant examples. For instance, *hune~bune* 'ship, boat' almost always appear as *bune* when it is non-initial in a word, but a few examples like *hiki+hune* 'tugboat' (cf. *hik-u* 'pull') deviate from the norm. Other inconsistent morphemes appear with rendaku in a relatively large fraction of relevant words and without rendaku in a relatively large fraction. For example, among common words ending in *ki~gi* (12a) the balance between those that show rendaku and those that do not is close to half and half.

12 Two of the immune morphemes in (11) have a medial /m/, and there was variability between /m/ and /b/ in many EMJ words, including those corresponding to *kemuri* and *himo* (Martin 1987: 31–32; Unger 2004: 331–332). The expectation is that Lyman's Law would have prevented rendaku in a morpheme containing a voiced obstruent, so it could be that immunity developed because of /b/ and then persisted even after the forms with /m/ eventually won out (Nakagawa 1966: 313–314). On the other hand, there are morphemes that showed this kind of variability historically but developed rendaku even so.

There are also inconsistent morphemes that only rarely show rendaku, one of which is *tuti~zuti* 'soil'. In fact, there are no common words that contain the voiced allomorph *zuti*, and it is tempting to classify this morpheme as rendaku-immune (Vance 1987: 147). But there is a famous kind of ceramics from the town of Imari in Saga Prefecture, and some present-day Tokyo speakers know the word *imari+zuti* 'clay used to make Imari ceramics'.

Rosen (2001: 35) refers to a native noun morpheme that typically appears with rendaku as a rendaku lover and to one that typically appears without rendaku as a rendaku hater. He treats rendaku-immune morphemes as a separate category (Rosen 2001: 40). Of course, these labels apply only to morphemes that begin with one of the voiceless obstruents in (1) when they occur word-initially, since rendaku would otherwise be impossible. Any morpheme containing a medial voiced obstruent is also beside the point, since Lyman's Law predicts its immunity to rendaku. The rendaku-immune morphemes of interest here are the ones that are immune for no apparent reason. Rosen makes some interesting claims based on small samples of compounds that he collected systematically using dictionaries. Most important, he says that in non-coordinate, two-element compounds in which both elements are native Japanese nouns and at least one of the two is three moras or longer, rendaku is predictable. (As shown below in section 5.2, coordinate compounds generally resist rendaku.) Rosen (2001: 70, 2003: 6) calls this generalization the prosodic size factor, but it seems more appropriate to refer to it as Rosen's Rule. To state the claim more explicitly, in a compound A+B that meets these criteria, as long as B begins with a voiceless obstruent as a word on its own and is not immune to rendaku, A+B will have rendaku. If this claim is correct, all native noun morphemes that are three moras or longer are either immune to rendaku or show rendaku consistently; the distinction between rendaku lovers and rendaku haters is relevant only to one-mora and two-mora elements. Furthermore, if A is three moras or longer and B is not rendaku-immune, then A+B will have rendaku regardless of whether B is a lover or a hater. Rosen's Rule is not in fact quite as ironclad as he suggests (Irwin 2009), but it is a very strong tendency even in a much broader range of vocabulary items, not restricted to two-element compounds containing only native elements.[13]

A careful look at the behavior of a few native noun morphemes will show how well Rosen's Rule seems to hold up. As mentioned earlier in this section, *hune~bune* 'ship, boat' is a rendaku lover but occasionally appears without rendaku. Rosen's Rule predicts that the voiceless allomorph *hune* will appear as the second element in a two- element compound only when the first element is shorter than three moras. The example cited above, *hiki+hune* 'tugboat', is in line with this prediction, although it is not on any of Rosen's lists because he restricted his search to noun+noun examples, and *hiki* is derived from the verb *hik-u* 'pull'. As a representative sample

13 There are some confusing inconsistencies in the way Rosen (2001) states his rule. The interpretation adopted here is arguably what he intended, but Irwin (2009) adopts a different interpretation.

of common words ending with *hune~bune*, we can use the relevant items that are listed both in a small reverse dictionary of Japanese (Kitahara 1990) and in a medium-size Japanese-English dictionary (Kondō and Takano 1986). There are 21 such words (including both headwords and compounds listed as sub-entries), nine with a first element longer than two moras, and these nine all have rendaku. Of the 12 with a shorter first element (one or two moras), eight have rendaku and four do not. In short, the 21 examples in this small sample conform to Rosen's Rule, even though the first elements are not restricted to single native noun morphemes. A much larger sample taken from a reverse-lookup dictionary (Sanseido Henshū-jo 1997) shows the same pattern (with one dubious exception), even though it includes many obscure and obsolete words that are not in the vocabulary of any ordinary speaker today.

As noted above, *ki~gi* 'tree; wood' (12a) is neither a lover nor a hater, since it shows rendaku in about half the relevant examples. Rosen's Rule predicts that only *gi* is possible in a compound with a first element longer than two moras, and a representative sample of common words from the same two dictionaries cited in the preceding paragraph (Kitahara 1990; Kondō and Takano 1986) conforms to this prediction. There are 38 words in the sample, and the six with a first element longer than two moras all have rendaku.[14] Of the 32 with a shorter first element, 13 have rendaku and 19 do not. Here again, the first elements are not restricted to single native noun morphemes. The much larger sample in the reverse-lookup dictionary (Sanseido Henshū-jo 1997) contains only one obscure compound that violates Rosen's Rule.

To test Rosen's Rule with a rendaku hater, compounds ending with *taka~daka* 'hawk' will serve. Only a small number of words fit this description, but there are enough to classify this morpheme tentatively as hater. In fact, only three examples are common enough to be listed in the dictionaries cited in the two preceding paragraphs (Kitahara 1990; Kondō and Takano 1986), and none of these has rendaku: *hage+taka* 'vulture' (cf. *hage-ru* 'go bald'), *kuma+taka* 'hawk eagle' (literally 'bear hawk'; cf. *kuma* 'bear'), and *yo+taka* 'nighthawk' (cf. *yo* 'night'). If the sample is expanded beyond very common words, however, there are examples containing the voiced allomorph. One is *aka+hara+daka* 'red-bellied hawk' (cf. *aka* 'red', *hara* 'belly'), which is listed in a field guide for birdwatchers (Takano 1982: 180) and denotes a species well known to people in this subculture. What is important for present purposes is that these examples with long first elements all have rendaku, so this rendaku hater conforms to Rosen's Rule.

14 The sample excludes reduplicated *ki+gi* 'trees' (see section 4.2) and coordinate *kusa+ki* 'grass and trees', which appears in (28) in section 5.2. The sample also excludes cases of folk etymology (e.g., *hiziki*, which denotes a type of seaweed but is sometimes written 肘木, as if it were a compound of *hizi* 'elbow' and *ki* 'tree'), and synchronically opaque etymological compounds or derivatives (e.g., *maki* 'Japanese yew', which is etymologically a combination of *ma* 'true' and *ki* 'tree').

Turning now to compounds with long final elements, as noted above, Rosen (2001: 70) restricted his claim to Japanese noun morphemes, so testing should begin by looking at native noun morphemes that are three moras or longer and begin with a voiceless obstruent as independent words. Rosen's Rule predicts that any non-coordinate compound ending with such a morpheme will have rendaku, unless the final element is rendaku-immune. It should not make any difference whether the immunity is idiosyncratic, as in the morphemes listed above in (11), or predictable by Lyman's Law (i.e., due to the inhibiting effect of a non-initial voiced obstruent). Rosen's generalization is that there are only two kinds of long native noun elements: those that always show rendaku (except in coordinate compounds) and those that never do. Rosen (2001: 232–237) lists examples ending in 40 different long elements, and his prediction holds up very well even if the sample is expanded to include all the obscure and obsolete compounds that end with one of these elements are listed in a large reverse-lookup dictionary (Sanseido Henshū-jo 1997). A few of Rosen's long elements are etymologically deverbal, including *tatami* 'tatami mat', which is related to the verb *tatam-u* 'fold up', but it is reasonable to treat all of these as nouns.

Among the long elements that Rosen (2001: 29) says are immune to rendaku, he is correct about *kanmuri* 'crown' and *kemuri* 'smoke', but not about *katati* 'shape'. The term *me+gatati* (cf. *me* 'eye') is used to denote a kind of position of the stones in the Japanese board game *go*, and it has rendaku. On the other hand, *kao+katati* 'features' (cf. *kao* 'face') and *mi+me+katati* 'looks' (cf. *mi-ru* 'see', *me* 'eye') are common enough words to be listed in Kondō and Takano (1986), and they lack rendaku. Rosen's Rule predicts that this kind of inconsistent behavior is impossible for a long element like *katati~gatati*. There are also a few inconsistencies among the long elements that Rosen says always show rendaku. Some of the small minority of relevant compounds that lack rendaku are obscure or obsolete, but two are common enough to be listed in Kondō and Takano (1986): *mi+sakai* 'distinction' (cf. *mi-ru* 'see', *sakai* 'boundary') and *yama+hutokoro* 'heart of the mountains' (cf. *yama* 'mountain', *hutokoro* 'bosom'). The rendaku in the common words *kuni+zakai* 'border between provinces' (cf. *kuni* 'province') and *uti+butokoro* 'real intentions' (cf. *uti* 'inside') seems to be more typical for these two long elements, although only a handful of compounds end with either one.[15] Nonetheless, the overall picture is clear: the overwhelming majority of noun+noun compounds with long final elements conform to Rosen's Rule. This is true even of Sino-Japanese elements (see section 4.3).

The explanation that Rosen (2003: 19–25) offers for his rule rests crucially on the notion of the foot. He assumes that foot boundaries must coincide with morpheme

15 Irwin (2005: 130) notes the contrast between *yama+hutokoro* and *uti+butokoro* and also points out the inconsistent behavior of three-mora *kitune~gitune* 'fox', which appears without rendaku in *kita+kitune* 'northern (red) fox' but with rendaku in all the other 23 compounds ending with this element that are listed in Sanseido Henshūjo (1997).

boundaries, and that a prosodic word contains at most two bimoraic feet. Consequently, a compound with a long first or second element is too big to fit into a single prosodic word, and the second element will therefore be at the beginning of a prosodic word. Rosen then argues that the “marked” [+voice] feature (i.e., rendaku) is permitted more freely in a prosodically strong position – the left edge of a prosodic word in this case. Leaving aside the mechanics of Rosen’s OT implementation of this idea, the question that arises, of course, is whether it is reasonable to claim that [+voice] is marked in the phonological environments in which rendaku is found.

An obstruent that shows rendaku is always preceded by a vowel or by the mora nasal /N/ and always followed by a vowel or by /y/.[16] The examples in (13) illustrate.

(13)	V – V	su+gao ‘face without makeup’ (cf. su ‘natural state’, kao ‘face’)
	/N/ – V	sin+gao ‘new face’ (cf. sin ‘new’)
	V – /y/	te+byoo·si ‘beating time’ (cf. te ‘hand’, hyoo·si ‘rhythm’)[17]
	/N/ – /y/	san+byoo·si ‘triple time’ (cf. san ‘three’)

Vowels, /N/, and /y/ are all typically voiced, so the obstruent that shows rendaku is more similar to its neighboring segments than a voiceless obstruent would be, although there is no phonotactic restriction against voiceless obstruents in any of the word-medial environments in (13) in modern Tokyo Japanese. A synchronic process or a diachronic change that voices voiceless obstruents in such environments is typically described as a phonetically motivated lenition, but Rosen’s account treats voicing as marked rather than unmarked when a word-medial obstruent is at the left edge of a prosodic word. For Rosen, rendaku is obligatory in a word like *makura+gi* ‘railroad tie’ (cf. *makura* ‘pillow’) because *ki~gi* ‘tree; wood’ is not rendaku-immune and there has to be a foot boundary conciding with the morpheme boundary. Since a prosodic word is limited to at most two feet, the long first element makes it impossible to incorporate the foot containing *gi* into the same prosodic word, and the result is $[_{\omega}[_{\varphi}\textit{maku}][_{\varphi}\textit{ra}]]+[_{\omega}[_{\varphi}\textit{gi}]]$. On the other hand, rendaku can be avoided in a word like *niwa+ki* ‘garden tree’ (cf. *niwa* ‘garden’) because the short first element makes it possible to incorporate the entire compound into a single prosodic word: $[_{\omega}[_{\varphi}\textit{niwa}][_{\varphi}\textit{ki}]]$. What is counterintuitive about Rosen’s account is that the stronger prosodic boundary between the morphemes in *makura+gi* triggers the allomorph of the second morpheme that is less natural in word-initial position.

The voiced/voiceless contrast operates word-initially in modern Tokyo Japanese, of course, so the rendaku alternations cannot be attributed to a phonotactic restriction against word-initial voiced obstruents, as the examples in (14) show.

16 This statement assumes a phonemic analysis that countenances syllable-initial C/y/ sequences like /by/, but it is possible to treat these same sequences as beginning with phonemically palatalized consonants like $/b^j/$ instead (Vance 2008: 226–232; Pintér, this volume).

17 The boundary between the morphs in *hyoo·si*, a Sino-Japanese binom (see section 4.3), is marked with a dot rather than a plus sign.

(14) a. initial /s/ ~ medial /s/
saki ‘tip’, tutu+saki ‘nozzle’ (cf. tutu ‘tube’)

b. initial /s/ ~ medial /z/
sao ‘pole’, take+zao ‘bamboo pole’ (cf. take ‘bamboo’)

c. initial /z/ ~ medial /z/
zeni ‘cash’, hi+zeni ‘daily cash income’ (cf. hi ‘day’)

Examples like (14a) and (14b) might suggest word-initial devoicing, but examples like (14c) show that the apparent pattern is not general. Nonetheless, the idea of analyzing the rendaku alternations as word-initial devoicing has some obvious appeal, and Kuroda (1963, 2002) explores how such an “anti-rendaku” analysis might be implemented formally. It is well known that voiced obstruents did not occur word-initially in Old Japanese (Takayama, this volume), except in the mimetic vocabulary and (possibly) borrowings from Chinese (Okumura 1972: 111; Martin 1987: 29–30). Even in modern Tokyo Japanese, very few native words begin a voiced obstruent, and there is little doubt that present-day speakers feel intuitively that word-initial voiced obstruents are marked in some sense (Martin 1987: 30). An initial devoicing analysis depends on finding some systematic way to identify and exclude vocabulary items that do not behave as it predicts, and this is no easy task.

4 Rendaku and vocabulary strata

4.1 Recent borrowings

It is well known that recent borrowings are generally immune to rendaku. That is, even if a recently borrowed morpheme is realized with an initial voiceless obstruent when it occurs as an independent word, it will not have an allomorph beginning with the paired voiced obstruent in (1). The examples in (15) illustrate with three recently borrowed morphemes (15a–c) and three comparable native morphemes (15d–f).

(15) a. kamera ‘camera’
i+kamera ‘gastro-camera’ (cf. i ‘stomach’)

b. tihusu ‘typhus’
tyoo+tihusu ‘intestinal typhus (cf. tyoo ‘intestine’)

c. hamu ‘ham’
nama+hamu ‘uncooked ham’ (cf. nama ‘raw’)

d. kame ‘turtle’
umi+game ‘sea turtle’ (cf. umi ‘sea’)

e. tikara ‘strength’
soko+zikara ‘underlying strength’ (cf. soko ‘bottom’)

f. hana ‘nose’
wasi+bana ‘hooked nose’ (cf. wasi ‘eagle’)

As an independent word, each of the native elements in (15d–f) begins with the same syllable as one of the recently borrowed elements in (15a–c). For each of these six morphemes, (15) gives one example of a compound containing that morpheme as its final element. Rendaku appears in each of the compounds ending with one of the native morphemes, but not in any of the compounds ending with one of the recently borrowed morphemes. Since there is no non-initial voiced obstruent in any of the recent loans, the absence of rendaku cannot be due to Lyman’s Law. This resistance to rendaku does not affect other elements that combine with recently borrowed elements. The examples in (16) are two-element compounds consisting of a recently borrowed first element and a native second element, and they all have rendaku.

(16) haato+gata ‘heart shape’ (cf. kata ‘shape’)
nekutai+dome ‘tie clip’ (cf. tome-ru ‘fasten’)
matti+bako ‘matchbox’ (cf. hako ‘box’)

A few morphemes borrowed from languages other than Chinese actually show rendaku, but these borrowings are all quite old. The examples in (17) illustrate. Both compounds are still in use and appear either as headwords or as examples even in small dictionaries.

(17) kappa ‘(rain) cape’ < Portuguese (earliest NKD citation: c. 1615)
ama+gappa ‘rain cape’ (cf. ama~ame ‘rain’)
karuta ‘(playing) cards’ < Portuguese (earliest NKD citation: 1596)
uta+garuta ‘poem cards’ (cf. uta ‘poem’)[18]

Kanji have been assigned to write the two loanword elements in (17). They can also be written in *katakana*, but it is not unusual to see them written in *hiragana* instead, and educated native speakers of Japanese are typically unsure about how to write them. This orthographic vacillation probably indicates uncertainty about the status

18 Younger speakers seem to be losing rendaku in compounds listed in dictionaries with *garuta* (Haruo Kubozono, p.c.).

of these loanword elements, since present-day norms prescribe *katakana* for non-Chinese borrowings. It is hard to say exactly how or when a loanword becomes so thoroughly integrated into the Japanese vocabulary that it starts behaving like a native or Sino-Japanese item.

Nakagawa (1966: 308) notes one sure example of rendaku in a non-Chinese loan that was borrowed much more recently than the elements in (17). Comprehensive dictionaries list both *ketto* ‘blanket’ (from English *blanket*; earliest NKD citation: 1872) and *aka+getto* ‘red blanket’, with the voiced allomorph of the borrowed element, although both words are obsolete. Nakagawa (1966: 308) also says that rendaku occasionally appears in examples such as *yama+gyanpu* ‘mountain camp’ (cf. *yama* ‘mountain’, *kyanpu* [from English] ‘camp’) and *indo+garee* ‘Indian curry’ (cf. *indo* ‘India’, *karee* [from English] ‘curry’), but he cautions that these pronunciations with rendaku are not yet established. In fact, rendaku in examples like these is no closer to general acceptance today, nearly 50 years later. Tokyo speakers typically react to such examples as jokes. A voiced allomorph of a non-Chinese loan is probably less likely to catch on today than in earlier periods, at least in part because the modern practice of writing such items in *katakana* helps to segregate them.

4.2 Mimetic elements and reduplication

Another sector of the Japanese vocabulary that resists rendaku is mimetic elements (see also Nasu, this volume). Many mimetic words are reduplicated, and such words do not show rendaku (Martin 1952: 49; Okumura 1955). The examples in (18) are typical.

(18)	kii+kii ‘screech-screech’	teku+teku ‘stride-stride’
	koso+koso ‘sneak-sneak’	ton+ton ‘tap-tap’
	siku+siku ‘sob-sob’	haki+haki ‘quick-quick’

Although nowhere near as abundant as reduplicated mimetic words like those in (18), non-reduplicated mimetic compounds also exist (Hamano 1998: 47–50). As the examples in (19) indicate, rendaku is also absent in these non-reduplicated compounds.

(19) petya+kutya ‘chitter-chatter’
cf. petya+petya ‘chatter-chatter’, kutya+kutya ‘chomp-chomp’
uro+tyoro ‘skitter-skatter’
cf. uro+uro ‘wander-wander’, tyoro+tyoro ‘flick-flick’

The pattern in (19) indicates that mimetic elements resist rendaku whether or not they are reduplicated. Non-reduplicated mimetic compounds are arguably coordinate,

however, and coordinate compounds generally resist rendaku (see section 5.2), so one could claim that the examples in (19) lack rendaku because they are coordinate, not because their second elements are mimetic. To demonstrate beyond doubt that mimetic elements are immune to rendaku, it would be necessary to find non-coordinate compounds with mimetic second elements. Hamano (1998: 55) lists four such examples, but the second elements all begin with /p/ or with a vowel, so none of the rendaku alternations in (1) is possible.

In sharp contrast to reduplicated mimetic words, rendaku is the norm in most other kinds of reduplicated words involving native Japanese bases. For instance, there are quite a few reduplicated words derived from a verb or an adjective. Words in this category strongly favor rendaku (provided Lyman's Law is not violated), even though they tend to be semantically and grammatically very similar to reduplicated mimetic words (Martin 1975: 410–411, 799–800). Some examples are given in (20).

(20) hore+bore 'fondly' (cf. hore-ru 'fall in love')
kasane+gasane 'repeatedly' (cf. kasane-ru 'repeat')
hoso+boso 'barely' (cf. hoso-i 'slender')
tika+zika 'soon' (cf. tika-i 'near')

There are even a few examples of rendaku in words that reduplicate the citation form of a verb (Martin 1975: 790–791), as in *kawaru+gawaru* 'by turns' (cf. *kawar-u* 'take the place of').

Reduplicated nouns also favor rendaku (respecting Lyman's Law, of course), as the examples in (21) show.[19]

(21) hito+bito 'people' (cf. hito 'person')
kona+gona 'smithereens' (cf. kona 'powder')
sore+zore 'each one' (cf. sore 'that one')
toki+doki 'sometimes' (cf. toki 'time')

Labrune (2012: 118) says that a reduplicated noun always show rendaku (as long as Lyman's Law is not violated) if the meaning is "plural or iterative" but not necessarily if the meaning is "distributive." The number of unambiguously distributive examples is small, and the only one in (21), *sore+zore*, has rendaku (although the NKD entry gives *sore+sore* as an alternative pronunciation).

19 There is extensive overlap between nouns and adverbs in Japanese, with many individual lexical items capable of functioning as either (Martin 1975: 782–817), and no attempt is made here to distinguish carefully between the two word classes for the examples cited. The distinction between ordinary nouns and adjectival nouns (*keiyō-dōshi*) is also ignored.

There are also a few adjectives containing a reduplicated base followed by the derivational suffix -*si*-, and these words too show rendaku in the reduplicated morpheme (unless it would violate Lyman's Law), as in (22).

(22) hana+bana+si-i 'splendid' (cf. hana 'flower')
karu+garu+si-i 'careless' (cf. karu-i 'light')

The reduplicated base in a word of this type is often hard to relate synchronically to any other existing vocabulary item, and even when the connection is obvious, the semantic relationship is often less than transparent. On the other hand, the word-formation pattern seems to be at least slightly productive synchronically. For example, semantically transparent *huyu+buyu+si-i* 'wintry' (cf. *huyu* 'winter') occurs in conversation, although it does not appear in dictionaries.

In contrast to the non-mimetic examples considered above, there is one type of native, non-mimetic reduplicated word that systematically resists rendaku. Reduplicating a verb base to convey the meaning 'while (repeatedly) doing' the action of the verb is a productive, though not frequently used, pattern in modern Japanese. Reduplications of this type are accentually unified, i.e., they are treated as single phonological words (Martin 1975: 408–409). For example, the verb *ka'k-u* 'write' yields *ka'ki+kaki* 'while writing'. Nonetheless, such words never show rendaku.

There is also a conspicuous resistance to rendaku in quasi-mimetic examples like *sima+sima* 'stripey' (on the label of a box of striped paper clips; cf. *sima* 'stripe') and *kani+kani* 'crab, crab, and more crab' (on posters advertising crab dinners; cf. *kani* 'crab'). Even recent borrowings can provide bases for quasi-mimetics, as in *rabu+rabu* 'lovey-dovey' (cf. *rabu* 'love' from English *love*). Rendaku is irrelevant in *rabu+rabu*, of course, since /r/ is not a voiceless obstruent, and a recent loan would be expected to resist rendaku in any case (see section 4.1), but some other explanation is required for *sima+sima* and *kani+kani*, since the bases are native nouns. Neither of these two morphemes is rendaku-immune, since the voiced allomorphs appear in examples like *yoko+zima* 'horizontal stripe' (1h) and *kabuto+gani* 'horseshoe crab' (cf. *kabuto* 'helmet').

It seems intuitively plausible to claim that examples like *sima+sima* 'stripey' resist rendaku because they are being treated as mimetic, and Nishimura (2013: 83–87) attempts to capture this intuition by distinguishing two kinds of reduplication. In "intensive/plural reduplication" the head is the base morpheme and appears on the right (reduplicant+base$_{H}$), and the reduplicated word inherits its syntactic category from the head. In contrast, "mimetic reduplication" yields words with "adjectival or adverbial meanings, even though the base stems are nouns" (Nishimura 2013: 85). The head is on the right in mimetic reduplication too, but it is the reduplicant rather than the base (base+reduplicant$_{H}$), so the head can be categorized as an adjective or an adverb, and the reduplicated word can inherit that category. This approach can successfully handle quasi-mimetic examples like *sima+sima*, but

the semantic distinction between the two kinds of reduplication is not clear-cut. Some of the well-established examples in (19) have adjectival or adverbial meanings, even though they show rendaku. It is also puzzling that reduplicated words derived from a verb or an adjective, like those in (20), favor rendaku so strongly, since (as noted) such words are semantically and grammatically very similar to reduplicated mimetic words. Nonetheless, the absence of rendaku in quasi-mimetic words may be the productive pattern today. Some verb-base reduplications that lack rendaku seem to belong in the quasi-mimetic category (Nishimura 2013: 98–100), including *suke-suke* 'see-through' (cf. *suke-ru* 'be transparent'), but examples like *kaki+kaki* 'while writing' (mentioned above) do not have quasi-mimetic adverbial or adjectival meanings, so the absence of rendaku remains unexplained.

4.3 Sino-Japanese elements and postnasal voicing

Sino-Japanese elements are much less likely than native Japanese elements to show rendaku. In fact, it is not unusual to encounter the claim that Sino-Japanese elements are immune to rendaku, but this claim is clearly false if we take it at face value and take Sino-Japanese elements to be morphemes that were adopted into Japanese in one of the three major waves of borrowing from Chinese.[20] The prototypical Sino-Japanese word is a binom, that is, a word written with two *kanji*, each *kanji* representing a Sino-Japanese morph. It is not difficult to find examples of rendaku affecting Sino-Japanese binoms, and a few are listed in (23). The boundary between the morphs in a Sino-Japanese binom is marked with a dot rather than a plus sign.

(23) waru+zi·e 'cunning' (cf. waru-i 'bad', ti·e 'wisdom')
ura+byoo·si 'back cover' (cf. ura 'back', hyoo·si 'cover')
boo·eki+gai·sya 'trading company' (cf. boo·eki 'trade', kai·sya 'company')
kaku+za·too 'cube sugar' (cf. kaku 'square', sa·too 'sugar')
mizu+dep·poo 'water pistol' (cf. mizu 'water', tep·poo 'gun')

In all the many examples of rendaku involving the initial consonant of a Sino-Japanese binom, there are no violations of Lyman's Law. Because of the limited variety of Sino-Japanese morph shapes, the only way a medial voiced obstruent can

20 The three waves of borrowing are known as *go-on* 'Wu pronunciations', *kan-on* 'Han pronunciations, and *tō-sō-on* 'Tang-Song pronunciations' (see Ito and Mester, this volume), the last of which was much smaller than the earlier two. It is often claimed that Sino-Japanese items (and items belonging to other strata) can be identified in terms of phonological behavior rather than etymology. Ito and Mester (1999) and Fukazawa and Kitahara (2005) provide relevant discussion. There are reasons for being skeptical about the notion of strata in general and also about this claim in particular (Vance 2002a; Ota 2004).

appear in a Sino-Japanese binom is as the first segment of the second morph, that is, right after the dot in the transcription.[21] The examples in (24) are typical.

(24) a. sya·kai+koo·zoo 'social structure'
cf. sya·kai 'society', koo·zoo 'structure'

b. aka+sin·goo 'red (traffic) light'
cf. aka-i 'red', sin·goo 'signal'

c. soo+hon·zan 'head temple'
cf. soo 'overall', hon·zan 'main temple'

These examples are potentially problematic if Lyman's Law has a morpheme as its domain, as theoretical treatments typically assume (see section 3.1). Using (24a) to illustrate, if a Sino-Japanese binom contains two morphemes, the /z/ in *koo·zoo* is not in the same morpheme as the /k/, so Lyman's Law should be irrelevant, but surely it is not just a coincidence that there are no examples like **sya·kai+goo·zoo*.

Although examples with rendaku like those in (23) are common, the great majority of Sino-Japanese binoms never show rendaku, even when Lyman's Law is irrelevant. In one comparison of a representative sample of 100 native Japanese monomorphemic nouns with a representative sample of 100 Sino-Japanese binoms (Vance 1996), 87 of the native elements showed rendaku in at least one compound, as opposed to only 10 of the Sino-Japanese elements. These numbers are just estimates, of course, but there is no doubt that rendaku is the norm for native Japanese noun morphemes, while immunity to rendaku is the norm for Sino-Japanese binoms.

In contrast to Sino-Japanese binoms as elements in longer words, the behavior of individual Sino-Japanese morphemes as elements within binoms raises a set of intractable problems related to rendaku that are too complex to go into here (Vance 1996, 2011). Sino-Japanese morphemes that occur alone (rather than as part of a binom) as final elements in compounds are less problematic. Irwin (2005) calls such elements mononoms and provides a thorough description of their rendaku behavior. A typical example is the last element in *tasi+zan* 'addition' (cf. *tas-u* 'add', *san* 'calculation').

One reason Sino-Japanese binoms are so problematic for a synchronic analysis of rendaku in modern Tokyo Japanese is that a process often called postnasal voicing (PNV) was active in Early Middle Japanese (800–1200). PNV left its mark mostly on the Sino-Japanese vocabulary, since nasal+obstruent sequences occurred mostly

21 If the name *saburoo* (see section 3.1) were categorized as a Sino-Japanese binom, it would require an amendment to this restriction on medial voiced obstruents and to the claim that rendaku affecting a binom never violates Lyman's Law (since it appears as *zaburoo* in longer names). Although *sabu* is etymologically a borrowing from Chinese, it is an irregular development from the source (cf. the regular development *san* 'three').

in Sino-Japanese words, but a subset of the reduction processes known as *onbin* 'euphonic changes' (see note 3) produced such sequences in some native words. The best-known native examples are verb forms like the modern Tokyo gerund *non-de* /noN-de/ (< $^{\mathrm{EMJ}}$*nomi-te*) 'drinking'. PNV was a neutralization, since its effect was that an obstruent immediately following a nasal had to be voiced. According to Frellesvig (2010: 307–308), PNV "ceased to apply as an automatic phonological rule" during the Late Middle Japanese period (1200–1600). Nonetheless, many phonologists assume that PNV is still active in modern Tokyo Japanese but that it applies exclusively or mainly to native Japanese elements (Ito and Mester 2003: 130–131; Tabata 2010: 98; Labrune 2012: 128–130). Ota (2004) and Rice (2005) offer persuasive counteraguments, but this is not the place to go into the details of the debate. Throughout this chapter, a voiced allomorph immediately following a nasal that looks like an instance of rendaku will be treated as just that: an instance of rendaku.

Rosen (2001: 28) says that he restricted the sample he used to test his generalization (see section 3.2) to noun+noun compounds in which both elements are of "Yamato origin" (i.e., native Japanese), but the etymological stratum of the elements in a compound seems to makes little if any difference for Rosen's Rule, although the matter needs to be investigated thoroughly. In particular, it looks as if ordinary Sino-Japanese binoms follow Rosen's Rule in most cases. Since all Sino-Japanese morphs are one or two moras long, a binom can range from two to four moras, so some are short (two moras), although most are long (three or four moras). Rosen's Rule says that a short final element that isn't rendaku-immune will always show rendaku after a long first-element in a non-coordinate compound. There is no indication that short second elements behave any differently when the first element is a Sino-Japanese binom as opposed to some other kind of item, although a systematic search for exceptions has not been carried out. For example, we saw above that *ki~gi* 'tree; wood' sometimes shows rendaku and sometimes does not, but as Rosen's generalization predicts, it appears as *gi* in *hyoo·si+gi* 'wooden clappers', which has the three-mora Sino-Japanese binom *hyoo·si* 'rhythm' as its first element. Rosen's Rule also says that a long second element in a non-coordinate compound will always show rendaku unless it is immune, and here again, it does not seem to matter whether the first element is a Sino-Japanese or something else. For example, three-mora *sakura~zakura* 'cherry tree/blossom' has rendaku in *hi·gan+zakura* 'cherry that blooms near the vernal equinox' (cf. the Sino-Japanese binom *hi·gan* 'equinoctial week'), just as it does in all other relevant examples.

As noted above, the great majority of Sino-Japanese binoms are immune to rendaku as final elements in longer words, but most of those that do show rendaku seem to show it consistently. The five binoms in (23) at least sometimes show rendaku, and one of them (*ti·e~zi·e* 'wisdom') is short, while the other four are long. This short binom always shows rendaku in a relevant compound, even when the first element is short, as in *saru+zi·e* 'shallow cunning' (cf. *saru* 'monkey'), although Rosen's Rule predicts only that *zie* should occur consistently after a long element.

For the long binoms, Rosen's Rule predicts consistent rendaku, and no relevant exceptions are listed in a large reverse-lookup dictionary (Sanseidō Henshū-jo 1997) for *sa·too* 'sugar'. The few exceptions listed for *hyoo·si* 'cover' and *kai·sya* 'company' are obscure or obsolete. The remaining binom in (23) is *tep·poo* 'gun', and the common word *mu+tep·poo* 'reckless' looks like an exception. Historically, *tep·poo* in *mu+tep·poo* is a folk etymology, but ordinary Tokyo speakers today think of it as containing the binom meaning 'gun', even though it is semantically opaque.[22] On the other hand, the Sino-Japanese quasi-prefix *bu* 'not' seems to inhibit rendaku in a following element (see section 5.5), and it could be that Sino-Japanese *mu* 'lacking' has a similar inhibiting effect. One binom that clearly violates Rosen's Rule is three-mora *sya·sin* 'photograph' (Irwin 2005: 135–136). It does not show rendaku in most relevant compounds, but the two exceptions are both common words: *ao+zya·sin* 'blueprint' (cf. *ao* 'blue') and *kao+zya·sin* 'mugshot' (cf. *kao* 'face'). This kind of inconsistency seems to be atypical, but only a thorough, systematic search can resolve the question.

Sino-Japanese mononoms are all short, since all Sino-Japanese morphs are one or two moras. If mononoms follow Rosen's Rule, they should always or never show rendaku following a long element, and *hon~bon* 'book' seems to conform to this prediction. Ohno (2000: 161) says that this element always has rendaku after a long element, as in *man·ga+bon* 'comic book' (cf. *man·ga* 'cartoon'), and almost never after a short element, as in *huru+hon* 'used book' (cf. *huru-i* 'old'). The only exceptions to Ohno's generalization involve *bon* following a short element, so these are not exceptions to Rosen's Rule. Whether or not mononoms in general follow Rosen's Rule remains to be investigated.

5 Rendaku and morphological/semantic structure

5.1 The right-branch condition

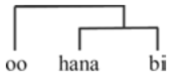
As mentioned in section 3.1, a constraint called the right-branch condition has been proposed to rule out rendaku in many compounds containing more than two elements. This constraint was first proposed by Otsu (1980: 217–222), and it says that rendaku can only appear in a morph that is on a right-side branch in the kind of branching diagram that shows semantic constituent structure.[23] The examples in (25) illustrate.

---

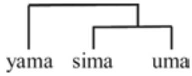
22 The NKD entry for *mu+tep·poo* gives two possible sources: *mu·te+hoo* 'empty-handed method' and *mu·ten+poo* 'no-mark method' (i.e., choosing not to annotate a Chinese text and thus leaving it open to misinterpretation).

23 Shibatani (1990: 175) suggests an alternative formulation involving the notion of a lexical head. For a comparison of the two versions of the constraint, see Vance (2007b).

(25) a. {oo+{hana+bi}} ‘grand fireworks’
cf. oo ‘big’, hana ‘flower’, hi ‘fire’
hana+bi ‘fireworks’



b. {yama+{sima+uma}} ‘mountain zebra’
cf. yama ‘mountain’, sima ‘stripe’, uma ‘horse’
sima+uma ‘zebra’



As (25a) shows, the three-element compound {*oo*+{*hana*+*bi*}} is a combination of *oo* ‘big’ and *hana*+*bi* ‘fireworks’ at the outer layer of compounding, and *hana*+*bi* contains a non-initial voiced obstruent, since its second element is the voiced allomorph *bi* of the morpheme meaning ‘fire’. If Lyman’s Law applies to each layer of compounding, as suggested above in section 3.1, the voiced allomorph *bana* of the morpheme meaning ‘flower’ would be a violation: *{*oo*+{*bana*+*bi*}}.

As (25b) shows, {*yama*+{*sima*+*uma*}} ‘mountain zebra’ has the same constituent structure as {*oo*+{*hana*+*bi*}}, but there is no voiced obstruent in *sima*+*uma*, so the absence of rendaku in *sima* cannot be attributed to Lyman’s Law. But *sima* is on a left-side branch in the diagram, so rendaku would violate the right-branch condition: *{*yama*+{*zima*+*uma*}}. Although some morphemes are immune to rendaku (see section 3.2), the morpheme meaning ‘stripe’ is not one of them, since the voiced allomorph *zima* appears common words such as *yoko*+*zima* ‘horizontal stripe’ (1h).

If Lyman’s Law is construed as an OCP effect that prohibits more than one voiced obstruent per morph (see section 3.1), it would not apply to (25a), since no morph in *{*oo*+{*bana*+*bi*}} contains more than one voiced obstruent. The right-branch condition predicts the absence of rendaku in not only in the middle element of {*yama*+{*sima*+*uma*}} (25b) but also in the middle elements of {*oo*+{*hana*+*bi*}} (25a) and {*hako*+{*hi*+*bati*}} ‘boxed brazier’ (9b), since each middle element is on a left branch. Ito and Mester (2003: 202–212) provide an OT analysis of the right-branch condition.

If the right-branch condition is a genuine constraint on rendaku, it predicts that the presence or absence of rendaku can sometimes serve to signal the constituent structure in compounds with more than two elements. Otsu (1980: 218–219) cites the two examples in (26) to make this point. Both contain the three elements *nuri* ‘lacquering’ (A), *hasi*~*basi* ‘chopsticks’ (B), and *ire* ‘putting in; container’ (C). Rendaku in B would violate the right branch condition if the constituent structure is {A{BC}} but not if the constituent structure is {{AB}C}.

(26) a. {{nuri+basi}+ire} ‘container for lacquered chopsticks’

b. {nuri+{hasi+ire}} ‘lacquered container for chopsticks’

In (26a) the voiced allomorph *basi* of the morpheme meaning ‘chopsticks’ is on a right branch, and in (26b) the voiceless allomorph *hasi* is on a left branch. Both

examples in (26) are novel compounds, but it does not seem at all implausible to suppose that a native speaker of Japanese could coin them by following productive morphological patterns. The important question to ask about such examples is whether the suggested relationship between rendaku and constituent structure corresponds to the intuitions of present-day Tokyo speakers. Some speakers (especially those who are linguists) have intuitions that are consistent with these examples, but it is not clear whether they are in the majority, and as Kozman (1998) reports, the responses of ordinary speakers on an experimental task also cast serious doubt on the psychological status of the right-branch condition. Presumably the right-branch condition is a genuine constraint for some speakers but not for others.

There are many apparent exceptions to the right-branch condition in the existing vocabulary. Those in (27), which are listed in a medium-size Japanese-English dictionary (Hasegawa et al. 1986), all begin with the bound element *oo* 'big'. As (27) shows, for each of these three-element words there is also an independent word consisting of the last two elements, so the constituent structures given for the three-element words are at least plausible.

(27) a. {oo+{date+mono}} 'star actor'
cf. tate-ru 'set up', mono 'person', tate+mono 'lead actor'

b. {oo+{buro+siki}} 'big wrapping cloth'
cf. huro 'bath', sik-u 'lay', huro+siki 'wrapping cloth'

c. {oo+{gane+moti}} 'very rich person'
cf. kane 'money', mot-u 'possess', kane+moti 'rich person'

There is no real doubt about the constituent structures for (27a) and (27b). In each case, the two-element compound that combines with *oo* is semantically opaque, but it is easy to analyze *tate+mono* and *huro+siki* into two elements each. Furthermore, there is no corresponding independent word *oo+date*, and while native speakers will accept *oo+buro* in the meaning 'big bath', {{*oo+buro*}+*siki*} is clearly wrong for the meaning 'big wrapping cloth'. The three-element word in (27c) is more problematic. In addition to *kane+moti*, *oo+gane* 'big money' exists as an independent word, and there is nothing strange about the structure {{*oo+gane*}+*moti*} for the meaning 'very rich person'. If *oo+gane+moti* is {{AB}C} instead of {A{BC}}, then it does not violate the right-branch condition.Some dictionaries mark the major division in a compound headword with a hyphen, and such dictionaries disagree about the constituent structure of *oo+gane+moti*.

Otsu (1980: 211–213, 220) tries to deal with apparent exceptions by distinguishing between what he calls loose and strict compounds. The basic idea is that a loose compound counts as two elements for the right-branch condition, whereas a strict compound counts as a single element. This proposal has some genuine intuitive

appeal, although it is probably impossible to draw a clear-cut distinction between strict and loose (Vance 1980a: 231–234). To illustrate with a simple example, using a superscript plus sign to mark the boundary between the elements of a strict compound, the claim would be that rendaku in *inu+go⁺ya* 'dog house' (cf. *inu* 'dog', *ko* 'small', *ya* 'house') does not violate the right-branch condition, since *ko⁺ya* 'hut' is a strict compound.

Sino-Japanese binoms that show rendaku, like those in (23), also violate the right-branch condition if they are treated as branching. For example, if *mizu+bu·soku* 'water shortage' (cf. *mizu* 'water', *hu·soku* 'insufficiency') is {*mizu*+{*bu·soku*}}, then *bu* (~*hu*) 'not' is on a left branch, but if *hu·soku* 'insufficiency' is like *ko⁺ya* 'hut', there is no violation. Sino-Japanese binoms are also problematic for Lyman's Law if its domain is a morph (see section 3.1), since rendaku never occurs in compounds like *tonari+kin·zyo* 'immediate neighborhood' (cf. *tonari* 'beside', *kin·zyo* 'neighborhood'), which ends with a Sino-Japanese binom that contains a voiced obstruent (**tonari+gin·zyo*). It looks as if Sino-Japanese binoms are indistinguishable from single morphemes both for the right-branch condition and for Lyman's Law.

5.2 Coordinate compounds

If a two-element compound A+B is a coordinate compound, A is not a modifier of B. Instead, the two elements have equal status, and the meaning of the compound can usually be paraphrased 'A and B'. It has been known for a long time that Japanese coordinate compounds resist rendaku (Lyman 1894: 9; Okumura 1955; Sakurai 1966: 41). The examples in (28) illustrate.

(28) oya+ko 'parent and child' (cf. osana+go 'young child', osana-i 'young')
kusa+ki 'grass and trees' (cf. nae+gi 'seedling tree', nae 'seedling')
tuki+hi 'days and months' (cf. naka+bi 'middle day', naka 'middle')

All the second elements in (28) show rendaku at least sometimes in non-coordinate compounds, and one example with rendaku is given for each second element. None of these second elements contains a non-initial voiced obstruent, so the absence of rendaku in the coordinate compounds cannot be attributed to Lyman's Law (cf. *migi+hidari* 'right and left').

A few coordinate compounds do show rendaku. Compounds containing an adjective element favor rendaku (see section 5.3), and Irwin (2012: 28) cites *ita+gayu-i* 'painful and itchy' (cf. *ita-i* 'painful', *kayu-i* 'itchy') and *ama+zuppa-i* 'sweet and sour' (cf. *ama-i* 'sweet', *suppa-i* 'sour') as coordinate. Irwin also notes *mie+gakure* 'appearing and disappearing' (cf. *mie-ru* 'become visible', *kakure-ru* 'become hidden'), and even though *mie+kakure*, without rendaku, exists as an alternative pronunciation, there are modern Tokyo speakers who accept only the form with rendaku as correct.

One other example is the three-element compound *asi+de+matoi* 'hindrance' (Nobue Suzuki, p.c.), which consists of *asi* 'foot', *te~de* 'hand', and *matoi* 'wrapping' (cf. *mato-u* 'wrap'). The figurative meaning 'hindrance' comes from the notion of binding a person's feet and hands, so the constituent structure is clearly {{*asi+de*}+*matoi*}, and despite the rendaku, the inner layer of compounding, *asi+de*, is obviously coordinate, even though there is no independent word *asi+de*. Although *asi+de+matoi* is more complex than the other examples cited, it is clearly relevant. Some dictionaries list *asi+te* (without rendaku) as a word meaning 'feet and hands', but *te+asi* 'hands and feet', with the two morphemes in the opposite order, is far more common. Many dictionaries give *asi+te+matoi* (without rendaku) as an alternative pronunciation, but there is no question that *asi+de+matoi* (with rendaku) is the current Tokyo norm (Shioda 2001: 101).

An even more complex example is *gen·kin+zi·doo+azuke+barai+ki* 'automated teller machine' (Wayne Lawrence, p.c.). The elements are the Sino-Japanese binoms *gen·kin* 'cash' and *zi·doo* 'automatic operation', the verb bases *azuke* (cf. *azuke-ru* 'entrust') and *harai~barai* (cf. *hara-u* 'pay'), and Sino-Japanese *ki* 'machine'. The constituent structure is {{*gen·kin*+{*zi·doo*+{*azuke+barai*}}}+*ki*}, and what is important for present purposes is that *azuke+barai* is clearly a constituent and clearly has the coordinate meaning 'depositing and repaying', presumably based on the two V+V=N compounds *azuke+ire* 'depositing' (cf. *ire-ru* 'put in') and *harai+modosi* 'paying back (as a withdrawal)' (cf. *modos-u* 'return').

5.3 Inflected words

Japanese has three classes of inflected words: verbs, adjectives, and the copula. Only verbs and adjectives participate in compounding, and rendaku could not affect the copula anyway, since the modern Tokyo forms begin with the voiced obstruent /d/. Okumura (1955: 962) claims that rendaku is unlikely in a two-element compound if both the elements are inflected words, but it is not immediately obvious exactly what this claim means (Vance 1987: 142–144). A reasonable interpretation is that if each root in a two-element compound is based on an inflected word, and the compound as a whole is an inflected word (i.e., a verb or an adjective), then rendaku is unexpected.

The examples in (29) are verb+verb compound verbs, that is, each compound is a verb and contains two verb roots. Words of this form are abundant in Japanese, and the abbreviation V+V=V is a convenient way to refer to them.[24]

24 There are good reasons for sub-categorizing V+V=V compounds into different types (Shibatani 1990: 246–247). Martin (1975: 438–439) distinguishes between compounds like those in (29) and cases where the second verb is what he calls an auxiliary. Kageyama (1999: 301–303) draws the same distinction and calls the two types lexical compound verbs and syntactic compound verbs. A V+V=V compound of the second type co-occurs with the same NPs as the initial element, has a com-

(29) kaki+tor-u ‘write down’ (cf. kak-u ‘write’, tor-u ‘take’)
oti+tuk-u ‘settle down’ (cf. oti-ru ‘fall’, tuk-u ‘arrive’)

The first component verb in a V+V=V compound has a constant form, and the second component verb takes whatever inflectional ending is required for the compound as a whole. In assessing the claim that rendaku is unlikely in such compounds, of course, only items that would otherwise allow rendaku are relevant. In particular, Lyman’s Law accounts for the absence of rendaku in an example like *nigiri+tubus-u* ‘squash by grasping’ (cf. *nigir-u* ‘grasp’, *tubus-u* ‘squash’). Even when such examples are excluded, V+V=V compounds seldom show rendaku. The standard account of the origin of rendaku sketched in section 2 provides a natural explanation for the rarity of rendaku in compounds of this type (Vance 1982: 340), since there is no reason to suppose that the two elements in a V+V=V compound were ever connected by a genitive particle or any other NV syllable in prehistoric Japanese.[25] Compound nouns containing two verb roots (V+V=N compounds) show rendaku far more often, and this difference has sometimes been exaggerated into the suggestion that verb/noun pairs like those in (30) are typical.

(30) V+V=V: ki+toos-u ‘wear continuously’
V+V=N: ki+doosi ‘continuous wearing’
cf. ki-ru ‘wear’, toos-u ‘make go through’
V+V=V: tukami+tor-u ‘take by grabbing’
V+V=N: tukami+dori ‘greedy snatching’
cf. tukam-u ‘grab’, tor-u ‘take’

Rendaku appears in both the compound nouns in (30) but not in either compound verb. Okumura (1955: 862) invites the inference that this pattern is typical pattern by citing a similar verb/noun pair as his only illustration.

In fact, however, pairs like those in (30) are not typical. The most common pattern by far is for both the verb and the noun in a pair to lack rendaku. There are also a few pairs that show rendaku both in the verb and in the noun. The examples in (31) illustrate these other two patterns.

---

pletely predictable meaning, and can be created on the spot rather than stored in the lexicon, since the pattern is productive. Also, as Kageyama (1999: 302–303) clearly explains, the two types show quite different behavior in a number of syntactic tests. Most of the V+V=V examples cited here are unmistakably the lexical type, but the distinction does not seem to be crucial here and is therefore ignored.

25 Some V+V=V compounds are coordinate, such as *tobi+hane-ru* ‘jump and leap’ (Tagashira and Hoff 1986: 7) and would be expected to resist rendaku for that reason (section 5.2).

(31) V+V=V: uti+kes-u 'negate' V+V=N: uti+kesi 'negation'
cf. ut-u 'strike', kes-u 'erase'
V+V=V: kaeri+zak-u 'reflower' V+V=N: kaeri+zaki 'reflowering'
cf. kaer-u 'return', sak-u 'bloom'

In a systematically collected sample of 234 relevant pairs (i.e., paired V+V=V and V+V=N compounds), 202 pairs (86.3%) do not have rendaku either in the V+V=V compound or in the V+V=N compound, 22 (9.4%) show the pattern in (30), and 10 (4.3%) show rendaku in both compounds (Vance 2005a: 93–98).

It is only unpaired examples that show a clear difference between V+V=V compounds and V+V=N compounds. An unpaired verb is a V+V=V compound with no coresponding noun. For example, there is no noun **okuri+kaesi* or **okuri+gaesi* corresponding to the verb *okuri+kaes-u* 'send back' (cf. *okur-u* 'send', *kaes-u* 'return'). (The inflectional form *okuri+kaes-i* is segmentally identical to the hypothetical noun but is not relevant here.) An unpaired noun is a V+V=N compound with no corresponding verb. For example, there is no verb **oboe+kak-u* or **oboe+gak-u* corresponding to the noun *oboe+gaki* 'memo' (cf. *oboe-ru* 'recall', *kak-u* 'write'). Among relevant unpaired verbs and nouns like these, rendaku occurs in a clear majority of the nouns but in only a tiny fraction of the verbs (Vance 2005a: 99).

Turning now to compounds containing adjective elements, a non-final adjective component has a constant form (identical to the root), and a word-final adjective component takes whatever inflectional ending is required for the compound as a whole. The abbreviation A+A=A refers to a compound that contains two adjective roots and is itself an adjective, e.g., *usu+gura-i* 'dimly lit' (cf. *usu-i* 'dim', *kura-i* 'dark'). There are also A+V=V compounds like *naga+bik-u* 'be prolonged' (cf. *naga-i* 'long', *hik-u* 'pull'), A+V=N compounds like *oso+zaki* 'late blooming' (cf. *oso-i* 'late', *sak-u* 'bloom'), and V+A=A compounds like *utagai+buka-i* 'suspicious' (cf. *utaga-u* 'doubt', *huka-i* 'deep'). A+A=N and V+A=N compounds are so rare that there is no point in trying to assess the likelihood of rendaku. One A+A=N example is *taka+hiku* 'unevenness' (literally 'highs and lows'; cf. *taka-i* 'high', *hiku-i* 'low'). Since this compound is coordinate, rendaku would be unexpected (see section 5.2). As for V+A=N compounds, all the apparent examples end with *daka* (cf. *taka-i* 'high'), as in *ure+daka* 'sales amount' (cf. *ure-ru* 'be sold'). Since the voiceless allomorph of this same adjective root occurs as the independent noun *taka* 'amount', it would be reasonable to analyze *ure+daka* as V+N=N rather than V+A=N.

All the examples cited in the preceding paragraph show rendaku, and rendaku seems to be the norm in relevant compounds involving adjective components, even when the compound itself is an adjective or a verb (Kikuda 1971; Vance 2005a: 98–99), although the number of examples of each type is small. Cases like V+A=A *mawari+kudo-i* 'roundabout' (cf. *mawar-u* 'go around', *kudo-i* 'wordy') are not relevant, of course, because Lyman's Law blocks rendaku. Coordinate examples like A+A=A *ama+kara-i* 'sugar and soy-sauce flavored' (cf. *ama-i* 'sweet', *kara-i* 'salty')

should probably be set aside as well, although two of the examples of rendaku cited in section 5.2 are A+A=A compounds (*ita+gayu-i* 'painful and itchy' and *ama+zuppa-i* 'sweet and sour'). Relevant A+A=A compounds divide about half and half into those with rendaku, as in *asa+guro-i* 'swarthy' (cf. *asa-i* 'shallow', *kuro-i* 'black'), and those without, as in *huru+kusa-i* 'old-fashioned' (cf. *huru-i* 'old', *kusa-i* 'smelly'). Relevant A+V=V compounds are very scarce, but they all show rendaku, as in *ara+date-ru* 'churn up' (cf. *ara-i* 'violent', *tate-ru* 'raise'). Relevant A+V=N compounds are more numerous, and most have rendaku, as in *waka+zini* 'premature death' (cf. *waka-i* 'young', *sin-u* 'die'). Most of the few relevant V+A=A compounds show rendaku. A typical example is *nebari+zuyo-i* 'tenacious' (cf. *nebar-u* 'stick', *tuyo-i* 'strong'), and one of the few exceptions is *tere+kusa-i* 'embarrassed' (cf. *tere-ru* 'get embarassed', *kusa-i* 'smelly').

N+A=N compounds like *ma+zika* 'close proximity' (cf. *ma* 'space', *tika-i* 'near') and N+A=A compounds like *ne+zuyoi* 'tenacious' (cf. *ne* 'root', *tuyoi* 'strong') also exist, but they have not been investigated carefully with respect to rendaku. In any case, they are beside the point here, since they contain an element that is not based on an inflected word.

To summarize, there does not seem to be any generalization that applies to all inflected-word compounds. In particular, the suggestion considered above in connection with V+V compounds does not work for compounds containing an adjective root. The suggestion was that rendaku is unlikely in a compound that meets two conditions: (1) the compound contains the roots of two inflected words, and (2) the compound itself is an inflected word. On the one hand, there is a real contrast between verbs and nouns containing two verb components: rendaku is rare in V+V=V compounds but common in V+V=N compounds. Rendaku is also common in all the compound types containing an adjective component, even when the compound as a whole is a verb (A+V=V) or an adjective (V+A=A or A+A=A). (As noted in section 4.2, reduplicated words based on a verb or adjective also strongly favor rendaku.) Incidentally, for compounds combining an adjective component with a verb component or with an another adjective component, the high frequency of rendaku is rather mysterious in terms of the explanation for the origin of rendaku offered in section 2. There is no compelling reason to think that some NV syllable would have appeared between the elements of such compounds in prehistoric Japanese.

5.4 Noun+verb compound nouns

A noun+verb compound noun contains a noun root followed by a verb root, as in *tiri+tori* 'dustpan' (cf. *tiri* 'dust', *tor-u* 'take'). Such N+V=N compounds are plentiful, and it has been proposed that rendaku is less likely if the noun element is in a direct-object relationship to the verb element (DO+V=N) rather than in some other

relationship (nonDO+V=N). Okumura (1955) and Sakurai (1966: 41) describe non-DO elements as adverbial modifiers, which presumably excludes subject elements. Examples involving the subject of the verb element are rare when the verb is transitive, as in *kami+kakusi* 'spriting away' (cf. *kami* 'god', *kakus-u* 'hide'), but not when the verb is intransitive. Kindaichi (1976: 12) suggests that Subject+V=N compounds resist rendaku regardless of whether the verb element is transitive or intransitive, but Sugioka (1986: 108, n. 24) disagrees. Subject elements will not be considered in the remainder of this section.

The claim that rendaku is less likely in DO+V=N compounds than in nonDO+V=N compounds is usually supported by citing examples that suggest a stronger generalization, namely, that nonDO+V=N compounds generally show rendaku, whereas DO+V=N compounds generally do not (Sugioka 2002: 500–501). If this stronger statement is correct, examples like those in (32) should be typical. Each example is accompanied by a noun+particle+verb phrase showing the semantic relationship between the noun component and the verb component, with direct objects marked by accusative *o* (as opposed to dative/locative *ni* or instrumental/locative *de*).

(32) DO+V=N: mono+hosi 'drying rack'
cf. mono o hos-u 'dry things'
nonDO+V=N: kage+bosi 'drying in the shade'
cf. kage de hos-u 'dry in the shade'
DO+V=N: kane+kasi 'money lender'
cf. kane o kas-u 'lend money'
nonDO+V=N: mae+gasi 'advancing money'
cf. mae ni kas-u 'lend in advance'
DO+V=N: azi+tuke 'flavoring'
cf. azi o tuke-ru 'put on flavor'
nonDO+V=N: kugi+zuke 'attaching with nails'
cf. kugi de tuke-ru 'attach with nails'

We see rendaku in the all three of the nonDO+V=N compounds in (32) (*kage+bosi*, *mae+gasi*, *kugi+zuke*) but not in any of the three DO+V=N compounds (*mono+hosi*, *kane+kasi*, *azi+tuke*). Compounds like *ude+kurabe* 'skill competition' (cf. *ude o kurabe-ru* 'compare skill') have to be set aside, of course, since a medial voiced obstruent like the /b/ in *kurabe* means that rendaku would violate Lyman's Law.

In fact, however, DO+V=N compounds with rendaku are common (Kindaichi 1976: 12–16). A few examples are listed in (33).

(33) hotaru+gari 'firefly hunting' (cf. hotaru o kar-u 'hunt fireflies')
hude+zukai 'brush technique' (cf. hude o tuka-u 'use a writing brush')
kuzi+biki 'drawing lots' (cf. kuzi o hik-u 'draw lots')

In one representative sample of common vocabulary items, about half the relevant DO+V=N compounds show rendaku (Nakamura and Vance 2002). As for nonDO+V=N compounds, the great majority of relevant items show rendaku, although not all. An example without rendaku is *kata+kake* 'shawl' (cf. *kata ni kake-ru* 'put on the shoulders'). In sum, it is true that rendaku is less common in DO+V=N compounds than in nonDO+V=N compounds, but it is not true that DO+V=N compounds strongly disfavor rendaku. At least in the established vocabulary, the difference between the two word types is that there is a very strong preference for rendaku in nonDO+V=N compounds but no clear preference either for or against in DO+V=N compounds. Kozman (1998) reports experimental results suggesting that there is no psychological reality to this tendency as a constraint on newly coined items, but Nakamura and Vance (2002) found that it seemed to be playing a role in a different experimental task. Sugioka (2005: 217–218) says that rendaku always occurs when it is possible in newly coined nonDO+V=N compounds but seldom occurs in newly coined DO+V=N compounds.

There is also a correlation between rendaku and accent in N+V=N compounds because DO+V=N compounds tend to be accented, whereas nonDO+V=N compounds tend to be unaccented (Sugioka 2002: 498–500; Yamaguchi 2011: 120). Consequently, the presence of an accent and the absence of rendaku tend to go together (in DO+V=N compounds), as do the absence of an accent and the presence of rendaku (in nonDO+V=N compounds). Akinaga (1966: 53) claims that this pattern holds only when the verb element is one or two moras long, but Yamaguchi (2011: 121–128), using a database of more than 1,000 relevant compounds listed in a dictionary, finds that the correlation is weaker but still significant when the verb element is three or four moras long. (Yamaguchi describes the first elements as arguments vs. adjuncts, but the argument-type examples in her database are all DO+V=N compounds). She also demonstrates that the probability of being accented is lower in N+V=N compounds that have rendaku, regardless of the relationship between the N and V components.[26]

Rosen's Rule (section 3.2) clearly does not hold in compounds that end with a verb element. It does not hold in V+V=V or V+V=N compounds, since they typically lack rendaku regardless of length (see section 5.3). It also does not hold in N+V=N compounds, as Rosen himself notes (Rosen 2001: 94), because of the inhibiting effect of the direct-object relationship.The direct-object relationship has a weaker but still statistically significant inhibiting effect when the verb element is long (three or four moras) as opposed to short (Yamaguchi 2011: 214). Sugioka (1986: 109–110) says that rendaku is common in N+V=V compounds like *ki+zuka-u* 'become concerned' (cf. *ki* 'mind', *tuka-u* 'use'), but words of this type have not been investigated systematically.

26 Sugito (1965) points out a similar tendency, confined to surnames ending with *ta~da* 'paddy', for rendaku and accent to be in complementary distribution. For a more comprehensive study of rendaku and accent, see Sato (1989).

5.5 Prefixes

All the preceding discussion indicates that the likelihood of rendaku in a two-element compound depends much more on the second element than on the first. The semantic relationship between the two elements is relevant in coordinate compounds (see section 5.2) and in N+V=N compounds (see section 5.4), and the length of the first element is relevant for Rosen's Rule (see section 3.2), but it is the phonological form of the second element that is relevant for Lyman's Law and the stratum of the second element (native, mimetic, Sino-Japanese, or recently borrowed) that correlates obviously with susceptibility to rendaku. The examples in (16) in section 4.1 were cited to show that recently borrowed first elements do not seem to inhibit rendaku. There are, however, some recent studies suggesting that first elements may have a subtle effect on the likelihood of rendaku responses in experimental tasks (Tamaoka et al. 2009; Tamaoka and Ikeda 2010).

As for prefix+base combinations, at least in some cases, it is clearly the first element that is important. Although the distinction between an affix and a bound root is not clear-cut (and no attempt will be made here to resolve this problem), the two honorific markers, native Japanese *o* and Sino-Japanese *go*, are uncontroversially prefixes, and a following base never shows rendaku.

Native *o* attaches mostly to native bases, as in *o+sake* 'rice wine', but there are also quite a few examples involving Sino-Japanese bases, as in *o+sa·too* 'sugar'. For examples like these to be relevant, of course, the second elements cannot be rendaku-immune, and *zi+zake* 'local rice wine' (cf. *zi* 'locality') shows that *sake~zake* alternates. Sino-Japanese elements are much less likely than native elements to show rendaku (see section 4.3), but one of the examples in (23) is *kaku+za·too* 'cube sugar' (cf. *kaku* 'cube'), so the binom *sa·too* is not immune. There are even a few examples of honorific *o* added to a recent loanword base, as in *o+soosu* 'sauce', but rendaku is unlikely in a recent borrowing under any circumstances (see section 4.1). The prefix *o* attaches not only to nouns but to adjectives, as in subject-exalting *o+tuyo-i* 'strong', and to the adverbial form of a verb in subject-exalting (honorific) and object-exalting (humble) constructions, as in subject-exalting *o+kaki ni nar-u* and object-exalting *o+kaki su-ru* (cf. the plain citation form *kak-u* 'write'). Rendaku never appears in any of these verb and adjective forms either.

Sino-Japanese honorific *go* attaches almost exclusively to Sino-Japanese noun bases. As noted in the preceding paragraph, some Sino-Japanese bases take the native prefix *o*, but of those that combine with an honorific prefix at all, most take *go*. Although the number of relevant examples is small, since only bases that are not rendaku-immune are relevant, it seems fair to say that *go*, just like *o*, blocks rendaku. The binom *ku·roo* 'hardship' shows rendaku in *ki+gu·roo* 'anxiety' (cf. *ki* 'mind'), but not in *go+ku·roo*, and there are no exceptions to this pattern.

Other first elements that seem to inhibit rendaku are native Japanese numerals (Nakagawa 1966: 314), especially *hito* 'one', as in *hito+koe* 'one cry' (cf. *koe* 'voice'),

and the Sino-Japanese quasi-prefix *bu* 'not' (Nakagawa 1966: 309–310), as in *bu+sai·ku* 'bungling' (cf. *sai·ku* 'craftsmanship'). Both these second elements typically show rendaku, as in *hana+goe* 'nasal voice' (cf. *hana* 'nose') and *tuno+zai·ku* 'horn carving' (cf. *tuno* 'horn'). As Irwin (2012: 31–32) shows, in contrast to native bare numerals, numbers, i.e., numeral+counter combinations like *hito+tubu* 'one grain; one drop', do not inhibit rendaku in the element that follows. For example, rendaku appears in *hito+tubu+dane* 'only child' (cf. *tane* 'seed').

5.6 Polysemy

Different senses of a polysemous morpheme often display markedly different behavior with respect to rendaku. To illustrate with just one such morpheme, *kuti~guti* (literally 'mouth') has a wide range of figurative meanings, and as the final element in a compound, the overall proportion of *guti* to *kuti* among frequently used words is roughly 2:1.[27] In the meaning 'doorway, gateway', it almost always shows rendaku, as in *ura+guti* 'back door' (cf. *ura* 'back') and *hi·zyoo+guti* 'emergency exit' (cf. *hi·zyoo* 'emergency'). On the other hand, in the meaning 'flavor', it consistently resists rendaku, as in *ato+kuti* 'aftertaste' (cf. *ato* 'after') and *ama+kuti* 'sweet taste' (cf. *ama-i* 'sweet'). When it comes to other senses of *kuti~guti*, there is less consistency, although in most cases compounds with *guti* are a clear majority. For example, for the sense 'speech, words', there are examples like *tuge+guti* 'tattling' (cf. *tuge-ru* 'tell') but also examples like *karu+kuti* 'jesting' (cf. *karu-i* 'light'), and there is variation in *waru+kuti~ waru+guti* 'bad mouthing' (cf. *waru-i* 'bad'). The existing vocabulary is full of narrowly circumscribed regularities and tendencies of this kind, and it seems very likely that ordinary speakers are sensitive to them, although probably not to the same degree as linguists. The picture is complicated by the fact that it is often hard to decide exactly which figurative meaning is involved, in part because (not surprisingly) the distinctions between different senses are not always clear-cut.

6 Unpredictability

6.1 Variation

Many vocabulary items exhibit variation between a form with rendaku and a form without. The examples in (34) are common enough words to be listed in NHK (1998)

27 This estimate of 2:1 for *guti* as opposed to *kuti* is based on compounds that appear as headwords in both the dictionaries described in section 3.2 (Kitahara 1990; Kondō and Takano 1986). The number of compounds involved is 69, and this total excludes *hito+kuti* (with a native numeral first element; see section 5.5), reduplicated *kuti-guti* (see section 4.2), and the frozen phrase *ko-i+kuti* 'strong flavor'. Several of the compounds are attested both with and without rendaku, although the dictionary entries usually do not reflect this variability (see section 6.1).

(a prescriptive pronunciation dictionary published by Japan's public broadcasting service, NHK). The alternative that appears on the left in each case is either the only pronunciation or the one given first in this dictionary. When the second alternative for a word in (34) is not given in the NHK dictionary, it appears either as the only pronunciation or as an alternative pronunciation for the corresponding headword either in Matsumura (2006) or in Shinmura (2008) (or in both).

(34) de+hune~de+bune 'sailing out' (cf. de-ru 'leave', hune 'boat')
koo·ri+kasi~koo·ri+gasi 'usury' (cf. koo·ri 'high interest', kas-u 'lend')
oku+huka-i~oku+buka-i 'deeply recessed' (cf. oku 'interior', huka-i 'deep')
sake+kuse~sake+guse 'behavior drunk' (cf. sake 'rice wine', kuse 'habit')
waru+kuti~waru+guti 'bad mouthing' (cf. waru-i 'bad', kuti 'mouth')
yaku+barai~yaku+harai 'exorcising evil' (cf. yaku 'evil', hara-u 'ward off')

It is not impossible that a single individual could sometimes use one form and sometimes use the other. More typically, however, a Tokyo speaker will use one form and regard the alternative form as mistaken or dialectal, although in some cases speakers will concede that the alternative form is acceptable.[28] Linguists, too, tend to underestimate the degree of variability, but Shioda (2001, 2011a, 2011b) has published some illuminating survey data in the monthly magazine put out by the NHK Broadcasting Culture Research Institute, which has been conducting surveys of fluctuations in the phonological form of words of since 1991.

6.2 Potential for disambiguation

Sometimes, although rarely, the presence or absence of rendaku corresponds to a difference in meaning, as in *oo+de* 'entire arm' vs. *oo+te* 'major company' (cf. *oo* 'big', *te* 'hand; arm'). (These two words also differ in accent for some Tokyo speakers: initial-accented *o'o+te* vs. unaccented *oo+de*, although some speakers have *o'o+de*). The presence vs. absence of rendaku in these two examples cannot be attributed to any difference in the pattern of combination. Since both are two-element compounds, the right-branch condition (see section 5.1) cannot be involved, and since neither is coordinate, the tendency for coordinate compounds to resist rendaku (see section 5.2) cannot be involved either.[29] The difference between *oo+de*

28 Although it is widely believed that traditional regional dialects differ significantly as far as rendaku is concerned, very little work has been done on this question (Vance, Miyashita, and Irwin, in press).

29 Linguists often cite coordinate *yama'+kawa* 'mountains and rivers' and non-coordinate (unaccented) *yama+gawa* 'mountain river', and the fact that *yama+gawa* does not have the coordinate meaning follows from a general pattern. Neither word is common enough to be listed in smaller dictionaries.

and *oo+te* is simply an example of the kind of semantic unpredictability that is-characteristic of compounding. While *oo+de* is semantically more transparent than *oo+te*, if it were not for the difference in pronunciation, a lexicographer might be tempted to treat both meanings as belonging to a single polysemous lexical item.[30] Compare English *handball*, which can denote either a game (in which the players hit a ball with their hands) or a soccer rule violation (i.e., touching the ball with a hand). Dictionaries typically list both meanings under the same headword, but it seems more reasonable to suppose that there are two different compounds consisting of *hand* and *ball*, presumably coined at different times in different places. The pronunciation difference between *oo+de* and *oo+te* compels lexicographers to treat them as two separate lexical items, but there is no principle behind the fact that one shows rendaku and the other does not. There is not even a tendency (of the sort described above in section 5.6) for the literal meaning 'hand' to favor rendaku. The meanings 'entire arm' and 'major company' could just as well be reversed, and this indeterminacy is symptomatic of how inconsistent rendaku is overall.

6.3 Analogy and the illusion of predictability

Despite all the tendencies cataloged above, as noted at the beginning of section 3.2, rendaku is fundamentally unpredictable. On the other hand, rendaku cannot possibly be a matter of memorizing which words have it and which words do not. Rendaku often occurs in newly coined words, and native speakers of Japanese who participate in experiments involving nonce words often produce or select responses with rendaku.

Ohno (2000: 161) proposes that a kind of analogy is the basic mechanism for extending rendaku to new vocabulary items. The idea is a native speaker accesses his/her lexicon for a semantically and/or phonologically parallel form and uses that form to decide whether or not a novel compound should have rendaku. For example, *kami~gami* 'hair' is a rendaku lover (see section 3.2), but it appears without rendaku in *kuro+kami* 'black hair'. When Ohno presented experimental participants with a novel compound written in *kanji* (白髪 'white'+'hair') and asked them to choose between *siro+kami* and *siro+gami* as the pronunciation, 27 of 31 chose *siro+kami*. Ohno's explanation is that the novel compound meaning 'white hair' strongly biases native speakers toward accessing the semantically parallel existing item *kuro+kami*. As a result, most experimental participants chose the form without rendaku for the

---

30 According to the relevant entries in NKD, early 13th-century words corresponding to modern Tokyo *oo+de* and *oo+te* are attested. The former already had its current meaning, but the latter meant 'front gate (of a castle)', and the modern meaning 'major company' developed from a longer word corresponding to modern Tokyo *oo+te+suzi*, which originally meant 'main road on the front side of a castle' (cf. *suzi* 'sinew', used figuratively to mean 'road') and then shifted by metonymy to mean 'major business'. Modern Tokyo *oo+te* is an abbreviation of this longer word.

novel compound. This is not the result that would be expected if the choice depended simply on the proportion of existing items with rendaku. As for the influence of phonologically parallel forms, Ohno compares the existing compounds *waka+kusa* 'young grass' (cf. *waka-i* 'young', *kusa* 'grass') and *i+gusa* 'rush' (cf. *i* 'rush'). In response to the novel compounds that Ohno used as test items, a majority (25/35) of the respondents preferred *aka+kusa* (without rendaku) over *aka+gusa* for 'red grass', but a majority (25/35) also preferred *ki+gusa* (with rendaku) over *ki+kusa* for 'yellow grass'. Ohno attributes the difference to the influence of phonologically similar *waka+kusa* and *i+gusa*.

Ohno (2000: 162) goes on to suggest that if there is no existing form to refer to, rendaku will not occur in a novel compound. If so, however, there is no explanation for why speakers extend rendaku to made-up elements in nonce-word experiments, as reported, e.g., in Vance (1980b), Ihara and Murata (2006), and Kawahara (2012) (see section 3.1). Despite this shortcoming, however, it seems likely that analogical decisions that turn on perceived similarity are an important factor in determining whether or not rendaku appears in an experimental response or in a newly coined vocabulary item. It seems plausible to suppose that when a particular individual is confronted with a particular novel item on a particular occasion, there is seldom any hesitation or doubt. As a result, rendaku in general feels predictable, even though different people do not always make the same choice and even though the same individual may make a different choice on a different occasion. It is not difficult to understand how his kind of feeling could translate into the widely-held (but clearly illusory) folk-belief that rendaku is regular.

There is a powerful temptation to claim that some apparent tendency is much more general than it actually is, and many clever amateurs are so convinced that rendaku is regular that they propose a new "rule" for every problematic vocabulary item. There are linguists, too, who seem unwilling to accept that rendaku is, to a significant degree, unpredictable. Searching tenaciously for heretofore undiscovered principles does no harm, of course, and it occasionally pays off, as in the discovery of Rosen's Rule (section 3.1). All indications are, however, that there is a hard core of intractable randomness in rendaku, and this is nothing out of the ordinary as far as morphophonemic alternations go.

Acknowledgments

The research reported here was supported in part by project funds from the National Institute for Japanese Language and Linguistics (NINJAL) and from the Research Institute for Humanity and Nature, by a travel grant from the Northeast Asia Council of the Association for Asian Studies, and by a grant-in-aid from the Japan Society for the Promotion of Science (no. 22820085). I would like to thank Mark Irwin, Shigeto

Kawahara, and Akio Nasu for reading and commenting perceptively on the first draft of this chapter. I am also grateful to all the members of the NINJAL Rendaku Project group for formal presentations and for informal conversations that have helped me to understand rendaku better.

References

Akinaga, Kazue. 1966. Kyōtsūgo no akusento [Accent of standard Japanese]. In Nihon Hōsō Kyōkai (ed.), *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary], 45–90. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Alderete, John. 1997. Dissimilation as local conjunction. In Kiyomi Kusumoto (ed.), *Proceedings of the North East Linguistic Society 27*, 17–32. Amherst: GLSA Publications.

Bloch, Bernard. 1946. Studies in colloquial Japanese I: Inflection. *Journal of the American Oriental Society* 66. 97–109.

Bunkachō. 1986. *Gendai kanazukai* (*naikaku kokuji*, *naikaku kunrei*) [Modern *kana* usage (cabinet notification and cabinet directive)]. Tokyo: Bunkachō.

Frellesvig, Bjarke. 2010. *A history of the Japanese language*. Cambridge: Cambridge University Press.

Frellesvig, Bjarke and John Whitman. 2008. Introduction. In Bjarke Frellesvig and John Whitman (eds.), *Proto-Japanese: Issues and prospects*, 1–9. Amsterdam: John Benjamins.

Fukazawa, Haruka and Mafuyu Kitahara. 2005. Ranking paradoxes in consonant voicing in Japanese. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 105–121. Berlin and New York: Mouton de Gruyter.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.

Hamada, Atsushi. 1952. Hatsuon to dakuon to no sōkansei no mondai [The correlation between moraic nasals and voiced obstruents]. *Kokugo Kokubun* 21(4). 18–32.

Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford/Tokyo: CSLI/Kurosio.

Hasegawa, Kiyoshi, Katsuaki Horiuchi, Chikara Momozawa and Saburō Yamamura (eds.). 1986. *Obunsha's comprehensive Japanese-English dictionary*. Tokyo: Obunsha.

Hashimoto, Shinkichi. 1932. Kokugo ni okeru biboin [Nasal vowels in Japanese]. *Hōgen* 2(1). 1–7. (References are to the version reprinted in Shinkichi Hashimoto. 1950. *Kokugo on'in no kenkyū* [Studies on Japanese phonology], 1–9. Tokyo: Iwanami.)

Hibiya, Junko. 1999. Variationist sociolinguistics. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 101–120. Oxford: Blackwell.

Hirano, Takanori. 1974. Rendaku no kisokusei to kigen [The regularity and origin of rendaku]. *Bungaku Kenkyū* 71. 21–43.

Ihara, Mutsuko and Tadao Murata. 2006. Nihongo no rendaku ni kansuru ikutsuka no jikken [Some experiments on sequential voicing]. *On'in Kenkyū* [Phonological Studies] 9. 17–24.

Irwin, Mark. 2005. Rendaku-based lexical hierarchies in Japanese: The behaviour of Sino-Japanese mononoms in hybrid noun compounds. *Journal of East Asian Linguistics* 14. 121–153.

Irwin, Mark. 2009. Prosodic size and rendaku immunity. *Journal of East Asian Linguistics* 18. 179–196.

Irwin, Mark. 2012. Rendaku dampening and prefixes. *NINJAL Research Papers* (*Kokuritsu Kokugo Kenkyūjo Ronshū*) 4. 27–36.

Irwin, Mark and Heiko Narrog. 2012. Late Middle Japanese. In Nicholas Tranter (ed.), *The languages of Japanese and Korea*, 246–267. Oxford: Routledge.

Ito, Junko and Ralf-Armin Mester. 1986. The phonology of voicing in Japanese. *Linguistic Inquiry* 17. 49–73.

Ito, Junko and Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Ms., Rutgers University (ROA-255-0498).

Ito, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge: MIT Press.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Iwanami Shoten Jiten Henshūbu (ed.). 1992. *Gyakubiki Kōjien* [Reverse-lookup *Kōjien*]. Tokyo: Iwanami.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kageyama, Taro. 1999. Word formation. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 297–325. Oxford: Blackwell.

Kawahara, Shigeto. 2012. Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment studies. *Lingua* 122. 1193–1206.

Kikuda, Norio. 1971. Yōgen no rendaku no ichi yōin [One factor affecting rendaku in inflected words]. *Kaishaku* 17(5). 24–29.

Kindaichi, Haruhiko. 1976. Rendaku no kai [On explaining rendaku]. *Sophia Linguistica* 2. 1–22.

Kitahara, Yasuo (ed.). 1990. *Nihongo gyakubiki jiten* [Reverse-lookup dictionary of Japanese]. Tokyo: Taishukan.

Klafehn, Terry. 2003. *Emergent properties of Japanese verb inflection*. Honolulu: University of Hawai'i dissertation.

Kondō, Ineko and Fumi Takano (eds.). 1986. *Shogakukan progressive Japanese-English dictionary*. Tokyo: Shogakukan.

Kozman, Tam. 1998. The psychological status of syntactic constraints on *rendaku*. In David Silva (ed.), *Japanese/Korean linguistics 8*, 107–120. Stanford: CSLI.

Kupchik, John. 2012. Morpheme-based *rendaku* as a rhythmic stabilizer in Eastern Old Japanese poetry. *Acta Linguistica Asiatica* 2. 9–22.

Kurisu, Kazutaka. 2001. *The phonology of morpheme realization*. Santa Cruz: University of California, Santa Cruz dissertation.

Kuroda, S.-Y. 1963. A historical remark on "rendaku," a phenomenon in Japanese morphology. Ms., MIT.

Kuroda, S.-Y. 2002. Rendaku. In Noriko M. Akatsuka and Susan Strauss (eds.), *Japanese/Korean linguistics 10*, 337–350. Stanford: CSLI.

Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Leben, William R. 2011. Autosegments. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume and Keren Rice (eds.), *The Blackwell companion to phonology 1*, 311–340. Malden, MA and Oxford: Wiley-Blackwell.

Lyman, Benjamin Smith. 1894. *The change from surd to sonant in Japanese compounds*. Philadelphia: Oriental Club of Philadelphia.

Maeda, Tomiyoshi. 1977. Rendaku. In Kiyoji Sato (ed.), *Kokugogaku kenkyū jiten*, 68. Tokyo: Meiji Shoin.

Maekawa, Kikuo. 2010. Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese. *Journal of Phonetics* 38. 360–374.

Martin, Samuel E. 1952. *Morphophonemics of standard colloquial Japanese*. (Supplement to *Language*, Language Dissertation No. 47). Baltimore: Waverly Press.

Martin, Samuel E. 1975. *A reference grammar of Japanese*. New Haven: Yale University Press.

Martin, Samuel E. 1987. *The Japanese language through time*. New Haven: Yale University Press.

Matsumura, Akira (ed.). 2006. *Daijirin* [Great dictionary], 3rd edition. Tokyo: Sanseido.

Mester, R. Armin and Junko Ito. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65. 258–293.

Miyake, Takeo. 1932. Dakuonkō [Thoughts on voiced obstruents]. *Onsei no Kenkyū* [The Study of Sounds, The Phonetic Society of Japan] 5. 135–190.

Murayama, Shichirō. 1954. Rendaku ni tsuite [On rendaku]. *Gengo Kenkyu* 26/27. 106–110.

Nakagawa, Yoshio. 1966. Rendaku, rensei (kashō) no keifu [The history of sequential voicing and (as a provisional term) sequential voicelessness]. *Kokugo Kokubun* 35(6). 302–314.

Nakamura, Kumiko and Timothy J. Vance. 2002. *Rendaku* in noun+verb-stem compounds: A production task. Paper presented at LP'2002, Meikai University, Urayasu, Chiba, Japan, 2–6 September.

Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Nihon Hōsō Kyōkai (ed.). 1998. *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Nishimura, Kohei. 2013. *Morphophonology in Japanese compounding*. Tokyo: University of Tokyo dissertation.

[NKD] Nihon Kokugo Daijiten Dainihan Henshū Iinkai (ed.). 2000–02. *Nihon kokugo daijiten* [Great dictionary of the Japanese language], 2nd edition. Tokyo: Shogakukan.

Odden, David. 2011. Rules vs. constraints. In John Goldsmith, Jason Riggle and Alan C. Yu (eds.), *The handbook of phonological theory*, 2nd edition, 1–39. Chichester: Wiley-Blackwell.

Ogura, Shinpei. 1910. Raiman-shi no rendakuron [Lyman's essay on rendaku]. *Kokugakuin Zasshi* 16 (7). 9–23; 16(8). 31–45.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer-Verlag.

Ohno, Kazutoshi. 2000. The lexical nature of rendaku in Japanese. In Mineharu Nakayama and Charles J. Quinn, Jr. (eds.), *Japanese/Korean linguistics 9*, 151–164. Stanford: CSLI.

Okumura, Mitsuo. 1955. Rendaku. In Kokugo Gakkai (ed.), *Kokugogaku jiten* [Glossary of Japanese linguistics], 961–962. Tokyo: Tōkyōdō Shuppan.

Okumura, Mitsuo. 1972. Kodai no on'in [Phonology of Ancient Japanese]. In Norio Nakata (ed.), *Kōza kokugoshi 2: On'inshi, mojishi* [Series in the history of Japanese 2: The history of phonology and orthography], 63–171. Tokyo: Taishukan.

Ota, Mitsuhiko. 2004. The learnability of a stratified lexicon. *Journal of Japanese Linguistics* 20. 19–40.

Otsu, Yukio. 1980. Some aspects of *rendaku* in Japanese and related problems. In Yukio Otsu and Anne Farmer (eds.), *Theoretical issues in Japanese linguistics* (MIT Working Papers in Linguistics 2), 207–227. Cambridge: MIT.

Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ramsey, S. Robert and J. Marshall Unger. 1972. Evidence for a consonant shift in 7th century Japanese. *Papers in Japanese Linguistics* 1. 278–295.

Rice, Keren. 2005. Sequential voicing, postnasal voicing, and Lyman's Law revisited. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 25–45. Berlin and New York: Mouton de Gruyter.

Rosen, Eric Robert. 2001. *Phonological processes interacting with the lexicon: Variable and non-regular effects in Japanese phonology*. Vancouver: University of British Columbia dissertation.

Rosen, Eric. 2003. Systematic irregularity in Japanese *rendaku*: How the grammar mediates patterned lexical exceptions. *Canadian Journal of Linguistics* 48. 1–37.

Sakurai, Shigeharu. 1966. Kyōtsūgo no hatsuon de chūi subeki kotogara [Matters requiring attention in the pronunciation of standard Japanese]. In Nihon Hōsō Kyōkai (ed.), *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary], 31–43. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Sanseido Henshūjo (ed.). 1997. *Kanjibiki gyakubiki daijirin* [Character-lookup and reverse-lookup *Daijirin*]. Tokyo: Sanseido.

Sato, Hirokazu. 1989. Fukugōgo ni okeru akusento kisoku to rendaku kisoku [Accent rules and rendaku rules in compound words]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 233–265. Tokyo: Meiji Shoin.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shinmura, Izuru (ed.). 2008. *Kōjien*, 6th edition. Tokyo: Iwanami.

Shioda, Takehiro. 2001. Yōgo no kettei [Usage decisions]. *Hōsō Kenkyū to Chōsa*, April, 96–105.

Shioda, Takehiro. 2011a. Yōgo no kettei [Usage decisions]. *Hōsō Kenkyū to Chōsa*, October, 98–105.

Shioda, Takehiro. 2011b. Yōgo no kettei [Usage decisions]. *Hōsō Kenkyū to Chōsa*, November, 86–97.

Sugioka, Yoko. 1986. *Interaction of derivational morphology and syntax in Japanese and English* (Outstanding Dissertations in Linguistics). New York: Garland Press.

Sugioka, Yoko. 2002. Incorporation vs. modification in deverbal compounds. In Noriko M. Akatsuka and Susan Strauss (eds.), *Japanese/Korean Linguistics 10*, 495–508. Stanford: CSLI.

Sugioka, Yoko. 2005. Multiple mechanisms underlying morphological productivity. In Salikoko S. Mufwene, Elaine J. Francis and Rebecca S. Wheeler (eds.), *Polymorphous linguistics: Jim McCawley's legacy*, 204–223. Cambridge: MIT Press.

Sugito, Miyoko. 1965. Shibata-san to Imada-san: Tango no chōkakuteki benbetsu ni tsuite no ichi-kōsatsu [Mr. Shibata and Mr. Imada: A remark on the perceptual discrimination of words]. *Gengo Seikatsu* 165. 64–72.

Suzuki, Yutaka. 2005. Raiman no hōsoku no reigai ni tsuite [On exceptions to Lyman's Law]. *Bunkyō Gakuin Tanki Daigaku Kiyō* 4. 249–265.

Suzuki, Yutaka. 2007. Raiman no nihongo kenkyū [Lyman's research on Japanese]. *Bunkyō Gakuin Daigaku Gaikokugo Gakubu Bunkyō Gakuin Tanki Daigaku Kiyō* 6 (2). 225–239.

Tabata, Toshiyuki. 2010. Sūshi "san" to "yon" ni tsuite [On the numerals *san* and *yon*]. In Hiroko Ōshima, Akiko Nakajima and Raoul Blin (eds.), *Kango no gengogaku*, 91–106. Tokyo: Kurosio.

Tagashira, Yoshiko and Jean Hoff. 1986. *Handbook of Japanese compound verbs*. Tokyo: Hokuseidō.

Takano, Shinji. 1982. *Fīrudo gaido: Nihon no yachō* [Field guide to the birds of Japan]. Tokyo: Nihon Yachō no Kai.

Takayama, Tomoaki. this volume. Chapter 15: Historical phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Tamaoka, Katsuo, Mutsuko Ihara, Tadao Murata and Hyunjung Lim. 2009. Effects of first-element phonological-length and etymological-type features on sequential voicing (rendaku) of second elements. *Journal of Japanese Linguistics* 25. 17–38.

Tamaoka, Katsuo and Fumiko Ikeda. 2010. Whiskey or bhiskey? Influence of first element and dialect region on sequential voicing of *shoochuu*. *Gengo Kenkyu* 137. 65–79.

Unger, J. Marshall. 1975. *Studies in early Japanese morphophonemics*. New Haven: Yale University dissertation.

Unger J. Marshall. 2004. Alternations of *m* and *b* in Early Middle Japanese: The deeper significance of the sound-symbolic stratum. *Japanese Language and Literature* 38. 323–337.

Vance, Timothy J. 1979. *Nonsense-word experiments in phonology and their application to* rendaku *in Japanese*. Chicago: University of Chicago dissertation.

Vance, Timothy J. 1980a. Comments on "Some aspects of *rendaku* in Japanese and related problems" by Yukio Otsu. In Yukio Otsu and Anne Farmer (eds.), *Theoretical issues in Japanese linguistics* (MIT Working Papers in Linguistics 2), 229–236. Cambridge: MIT.

Vance, Timothy J. 1980b. The psychological status of a constraint on Japanese consonant alternation. *Linguistics* 18. 245–267.

Vance, Timothy J. 1982. On the origin of voicing alteration in Japanese consonants. *Journal of the American Oriental Society* 102. 333–341.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: SUNY Press.

Vance, Timothy J. 1991. A new experimental study of Japanese verb morphology. *Journal of Japanese Linguistics* 13. 145–166.

Vance, Timothy J. 1996. Sequential voicing and Sino-Japanese. *Journal of the Association of Teachers of Japanese* 30. 22–43.

Vance, Timothy J. 2002a. Another look at vocabulary stratification in Japanese. In Fabrice Cavoto (ed.), *The linguist's linguist: A collection of papers in honour of Alexis Manaster Ramer*, 433–436. Munich: Lincom Europa.

Vance, Timothy J. 2005a. Rendaku in inflected words. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 89–103. Berlin and New York: Mouton de Gruyter.

Vance, Timothy J. 2005b. Sequential voicing and Lyman's Law in Old Japanese. In Salikoko S. Mufwene, Elaine J. Francis and Rebecca S. Wheeler (eds.), *Polymorphous linguistics: Jim McCawley's legacy*, 27–43. Cambridge: MIT Press.

Vance, Timothy J. 2007a. Have we learned anything about *rendaku* that Lyman didn't already know? In Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith (eds.), *Current issues in the history and structure of Japanese*, 153–170. Tokyo: Kurosio.

Vance, Timothy J. 2007b. Right branch or head: What difference does it make? In Susumu Kuno, Seiichi Makino and Susan Strauss (eds.), *Aspects of Japanese linguistics*, 221–240. Tokyo: Kurosio.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.

Vance, Timothy J. 2011. *Rendaku* in Sino-Japanese: Reduplication and coordination. In Ho-Min Sohn, Haruko Minegishi Cook, William O'Grady, Leon A. Serafim and Sang Yee Cheon (eds.), *Japanese/Korean linguistics 19*, 465–482. Stanford: CSLI.

Vance, Timothy J. 2012. Benjamin Smith Lyman as a phonetician. *Journal of Japanese Linguistics* 28. 31–41.

Vance, Timothy J., Mizuki Miyashita and Mark Irwin. In press. *Rendaku* in Japanese dialects that retain prenasalization. In Seungho Nam, Heejeong Ko and Jongho Jun (eds.), *Japanese/Korean Linguistics 21.* Stanford: CSLI.

Yamaguchi, Kyoko. 2011. Accentedness and *rendaku* in Japanese deverbal compounds. *Gengo Kenkyu* 140. 117–134.

Yoshida, Sumio and Yūichi Inokuchi (eds.). 1962. *Meiji ikō kokuji mondai shoan shūsei* [Complilation of proposals on orthography since the Meiji era]. Tokyo: Fuzama Shobō.

IV Prosody

Shigeto Kawahara

11 The phonology of Japanese accent

1 Introduction

1.1 Background and the aims of this chapter

The Tokyo dialect of Japanese exhibits lexical contrasts based on pitch accent; that is, there are minimal pairs of words that are identical segment-wise,[1] but can be distinguished in terms of their pitch contours (the term "accent" is defined shortly below in section 1.2). While what kind of pitch contour a particular word shows is often unpredictable for many lexical words, there are many phonological and morphological environments in which the distribution of lexical accent is predictable, at least to some extent. In other words, there are some regularities regarding the phonological distributions of Japanese pitch accent. This chapter provides an overview of the phonology of pitch accent patterns in modern Tokyo Japanese (henceforth "Japanese").

Since the accentual system of Japanese is so complex, it is impossible to provide a full description of its system, let alone an analysis, in a single chapter. Many details of Japanese accentology therefore have to be set aside. For example, although there is a wealth of literature on the accent patterns of non-Tokyo dialects, it is far beyond the scope of this chapter to discuss them. See, for example, Haraguchi (1977, 1991, 1999), Kubozono (2010, 2011), and Uwano (1999, 2007) for some descriptions of non-Tokyo dialects written in English. Neither does this chapter go into the details of phonetic realization of Japanese accent (for which see Beckman 1986, Pierrehumbert and Beckman 1988, Poser 1984 and Sugiyama 2012 and references cited therein, as well as Igarashi, this volume, and Ishihara, this volume). This chapter instead provides an overview of the complex patterns of Tokyo Japanese accentology with an emphasis on the description of the system, while also discussing it from the cross-linguistic perspective of metrical phenomena in other languages.

The aim of the current chapter is to make the materials accessible to those who have little or no knowledge of Japanese phonetics and phonology, although this chapter does assume some familiarity with basic phonological notions in some parts of the discussion. Readers are also referred to other overview articles (Akinaga 1985; Haraguchi 1999; Kubozono 2008, 2011, 2013) and relevant chapters on accent in

---

1 Presence of accent does affect the phonetic realization of segments in dimensions other than fundamental frequency; for example, accented syllables are slightly longer than unaccented syllables (Hoequist 1982). See Beckman (1986), Pierrehumbert and Beckman (1988), Poser (1984), and Sugiyama (2012) and references cited therein, as well as some discussion in Igarashi (this volume) and Ishihara (this volume) for the phonetics of Japanese pitch accent.

books on Japanese phonology (Labrune 2012; Vance 1987, 2008) for further discussion and references, although this chapter itself draws heavily on them.

The rest of this chapter proceeds as follows. The remainder of this introduction clarifies the terms and introduces the basic phonetic and phonological nature of Japanese pitch accent. Section 2 discusses accent patterns of loanwords, which have been argued to reflect the default accent assignment rule in Japanese. Section 3 observes that the default pattern may be reflected in the Japanese lexicon in a stochastic way. Section 4 is a discussion of compound accent rules, which have attracted much attention in the literature. Section 5 briefly provides an overview of the accent patterns of verbs and adjectives. Section 6 discusses several types of affixal accent patterns. Section 7 presents some other domains of Japanese phonology in which accent patterns are more or less predictable. Section 8 discusses how accent patterns interact with other phonological patterns in Japanese. Section 9 presents some remaining issues, and Section 10 is an overall conclusion.

1.2 Clarification of the terms used

To begin our discussion, some clarification of the term "pitch accent" may be useful. There are two senses in which the term "pitch accent" can be and has been used in the literature. A pitch accent can refer to an abrupt fall in fundamental frequency (i.e., F0 or pitch[2]) that is found in many words in Tokyo Japanese; for example, one finds a statement like "the word /kokoro/ 'heart' has pitch accent on the second syllable".[3] When the term is used in this sense, it refers to a physical, acoustic event, that is, a tonal fall found from the second syllable to the third syllable, or it can refer to phonological prominence associated with that tonal fall.

The same term "pitch accent" can also refer to a lexical contrast based on the presence or location of that pitch fall; when the term is used in this sense, it refers to a phonological distinction or property. For instance, we can talk about "the accent of loanwords", "the accent of adjectives", or even "the accent of unaccented words". See the Introduction to this volume for more on the ambiguity of this term. Finally, the term "pitch accent" does not refer here to – as it would in describing languages like English (Bolinger 1958) – phrasal prominence that is assigned to focused constituents. Pitch accent in Japanese is fundamentally a word-level property, not a phrasal or sentence-level property, although it interacts non-trivially with sentence-level intonational patterns (see Igarashi, this volume, and Ishihara, this volume for more on the interaction between word-level accent and sentence-level tones).

2 The term "pitch" is sometimes used to refer to a perceptual correlate of F0 (fundamental frequency), which is on the other hand an acoustic/physical property – how many times the glottis vibrates per second. It is common, however, in the Japanese literature to use the term "pitch" to refer to the acoustic event (fall in F0) rather than the perceptual property, and this chapter follows that convention.

3 For the sake of simplicity, examples in this chapter are given in romanized phonemic forms rather than phonetic transcriptions.

1.3 Pitch contrasts in Japanese

Having clarified the meanings of the term “pitch accent”, we now turn to how Japanese accent is mapped onto actual tonal (or F0) patterns. First, setting aside the precise phonetic realizations, Japanese makes lexical contrasts in terms of pitch accent in two ways: (i) presence vs. absence, and (ii) if present, location. The examples in (1) illustrate the lexical contrast based on the presence vs. absence of pitch accent.[4]

(1) Minimal pairs of unaccented and accented words

a.	ame+ga (unaccented)	‘candy+NOM’
b.	a’me+ga (accented)	‘rain+NOM’
c.	sake+ga (unaccented)	‘alcohol+NOM’
d.	sa’ke+ga (accented)	‘salmon+NOM’
e.	kaki+ga (unaccented)	‘persimmon+NOM’
f.	ka’ki+ga (accented)	‘oyster+NOM’
g.	kaku+ga (unaccented)	‘rank+NOM’
h.	ka’ku+ga (accented)	‘core+NOM’
i.	aki+ga (unaccented)	‘availability+NOM’
j.	a’ki+ga (accented)	‘autumn+NOM’

Whereas the words in (1a, c, e, g, i) are unaccented, those in (1b, d, f, h, j) are accented. It is common to represent the presence and location of accent with /’/ after the accented syllable. Phonetically speaking, an accented vowel is assigned a High tone followed by a Low tone on the following vowel, resulting in an abrupt H(igh)-L(ow) fall in F0, whereas unaccented words do not show such a fall. The use of this diacritic /’/ has the virtue of directly representing this phonetic implementation of Japanese pitch accent. Unlike in many other tonal languages (Yip 2002), Japanese lexically uses only two levels of tonal heights (High and Low, and not, for example, Mid).[5]

---

4 A few notes about data presentation and data sources in this chapter are in order. This chapter uses the following conventions to denote several types of boundaries: “+” for morphological boundaries; “-” for mora boundaries; and “.” for syllable boundaries – see Kubozono’s introduction to this volume and Otake (this volume) for the nature of the moraic system in Japanese. In illustrative examples, the nominative marker /+ga/ is often attached – the reason for this convention will become clear shortly. The data in this chapter come from various sources cited below, including the NHK dictionary (NHK 1998), as well as from suggestions from my colleagues; there are cases in which the accent locations are based on the author’s intuition as a native speaker of Tokyo Japanese. This intuition-based approach may not be the optimal methodology for data collection in linguistics, but this approach is deployed for practical reasons in this chapter. See section 10.1 for some discussion.

5 McCawley (1968) used Mid to represent downstepped H, a lowered H tone following another H tone (see Igarashi, this volume, and Ishihara, this volume). Complex tonal interactions occur at phrasal and sentential levels, which, phonetically speaking, result in many more than binary tonal height (Pierrehumbert and Beckman 1988, Kawahara and Shinya 2008, and Igarashi, this volume, and Ishihara, this volume); however, at the lexical level, it is safe to say that Japanese makes use of only two level tones.

Japanese also distinguishes words in terms of where pitch falls; i.e., in terms of accent location. This contrast in accent placement is exemplified in (2), where the words in (a, c, e) are accented on their initial syllables, while the words in (b, d, f) have final accent. A classical set of examples showing the "*n*+1 pattern" (Akinaga 1985; Haraguchi 1999; McCawley 1968; Shibatani 1990; Uwano 1999, 2007) is given in (3), where for words consisting of *n*-syllables, there are *n*+1 accent patterns (McCawley 1968: 138). In this particular case, for trisyllabic words, we find four distinct accent patterns: accent can fall on any of the *n*-th syllables, and there can additionally be an unaccented word.

(2) Minimal pairs illustrating the contrastiveness of accent locations
  a. ka'ta+ga (initial accent) 'shoulder+NOM'
  b. kata'+ga (final accent) 'frame+NOM'
  c. ko'to+ga (initial accent) 'Japanese zither+NOM'
  d. koto'+ga (final accent) 'matter+NOM'
  e. ka'ki+ga (initial accent) 'oyster+NOM'
  f. kaki'+ga (final accent) 'fence+NOM'

(3) *n*+1 accent pattern
  a. i'noti+ga (initial accent) 'life+NOM'
  b. koko'ro+ga (penultimate accent) 'heart+NOM'
  c. atama'+ga (final accent) 'head+NOM'
  d. miyako+ga (unaccented) 'city+NOM'

According to Sibata and Shibata (1990), cited by Kubozono (2001a) and Labrune (2012), 14% of minimal pairs in Japanese are distinguished by a pitch contrast.

A few final remarks are in order. First, although the Tokyo dialect of Japanese allows *n*+1 accent patterns, this description does not hold for words of any syllable length. Especially in long words (words longer than 4 moras, in particular), words with initial or final accent are rare at best (Kawahara and Kao 2012; Kubozono 2008; Labrune 2012; Sibata 1994).

Second, there is a non-negligible degree of inter-speaker as well as intra-speaker variability in accent placement. For example, the word for 'cousin' can be pronounced as /i'toko/ (with initial accent) or /ito'ko/ (with penultimate accent). The word for 'mind' can be /koko'ro/ (with penultimate accent) or /kokoro'/ (with final accent). In some cases, different accent assignments may be due to the influence of non-Tokyo dialects. The data presented in this chapter, therefore, involves some level of simplification and abstraction by the author, and not every speaker of Tokyo Japanese may agree with all the data presented here.

1.4 From pitch accent to surface tones

Now we turn to how these accent patterns are mapped onto surface tonal patterns. A HL fall in F0 occurs across the two syllables separated by /'/; in other words, the

accented vowel bears a H tone and the following vowel bears a L tone, as schematically illustrated in (4).

(4) Tones assigned by accent

a. ka’ta+ga ‘shoulder+NOM’

```
k  a’  t  a  g  a
   |      |
   H      L
```

b. kata’+ga ‘frame+NOM’

```
k  a  t  a’  g  a
         |      |
         H      L
```

Aside from the tones assigned by pitch accent, the first two syllables in a word bear a LH tonal sequence, sometimes known as initial lowering or initial rise, unless the first syllable is accented.[6] (5a) illustrates the tonal assignment due to initial rise for the word /atama’+ga/ ‘head’. When the initial syllable is accented, the word receives the accentual HL fall instead; i.e., initial rise does not apply, as in (5b).

(5) Tones assigned by initial rise

a. atama’+ga ‘head+NOM’

```
a  t  a  m  a’  g  a
|     |     |      |
L     H     H      L
```

b. ka’ta+ga ‘shoulder+NOM’ (Initial rise is blocked by initial accent)

```
k  a’  t  a  g  a
   |      |
   H      L
```

Finally, when syllables do not receive a tonal specification either from pitch accent or from initial rise, then these tonally-unspecified syllables get their tonal specifications by copying the tone from the rightmost specified syllable, which results in the forms like (6). This term “copying” is used here as a descriptive term; Haraguchi (1977), for example, achieves this result by autosegmental spreading (Goldsmith 1976), the notation which is used in (6). It may alternatively be better

6 Some researchers consider this initial lowering as a case of tonal dissimilation (Haraguchi 1977, 1991, 1999; Labrune 2012), whereas others, including the J-ToBI transcription system (Maekawa et al. 2002; Venditti 2005), consider the initial L tone to be a phrasal tone (Kawakami 1961; Pierrehumbert and Beckman 1988). See also Igarashi (this volume) and Ishihara (this volume). When the initial syllables contain a long vowel (e.g., /tookyoo/ ‘Tokyo’), they can be pronounced with HH without initial lowering (Haraguchi 1977, 1991; Vance 1987). See again Igarashi (this volume).

characterized as phonetic interpolation, in which case syllables that do not receive tones either from pitch accent or initial rise are toneless phonologically even at the surface level (Pierrehumbert and Beckman 1988 – see Igarashi, this volume, for further discussion and Myers 1998 for more on tonal phonetic underspecification at the surface level).

(6) Tones assigned by tonal spreading/copying/interpolation

a. ka'ta+ga 'shoulder+NOM'

k a' t a g a

H L

b. ame+ga 'candy+NOM'

a m e g a

L H

As a result of these tonal assignment mechanisms, all syllables receive tonal specifications. For example, initially-accented trisyllabic words receive a HLL tonal contour, whereas medially-accented trisyllablic words receive a LHL tonal contour.

To summarize, the tonal shape of a particular word can be completely determined by the presence/absence of a pitch accent and its location. The derivations in (7–9) illustrate how each accent pattern receives its full tonal specification, taking unaccented, initially-accented, and medially-accented words as examples.[7]

(7) From accent to tones: Unaccented nouns

1. Underlying form

x x x x x

2. Accentual tone assignment (does not apply)

x x x x x

3. Initial rise

x x x x x

L H

4. Tonal spreading

x x x x x

L H

7 This model is just an example. For various proposals on how to represent Japanese accent underlyingly and how to derive surface tonal patterns from particular underlying representations, see Haraguchi (1977), Pierrehumbert and Beckman (1988), Poser (1984), and Pulleyblank (1984).

(8) From accent to tones: Initial accent

1. Underlying form

x' x x x x

2. Accentual tone assignment

x' x x x x

H L

3. Initial rise (does not apply)

x' x x x x

H L

4. Tonal spreading

x' x x x x

H L

(9) From accent to tones: Antepenultimate accent

1. Underlying form

x x x' x x

2. Accent assignment

x x x' x x

H L

3. Initial rise

x x x' x x

L H H L

4. Tonal spreading

x x x' x x

L H H L

Since accent is realized as a HL fall, the distinction between finally-accented words (e.g., /kaki'/ 'fence') and unaccented words (e.g., /kaki/ 'persimmon') are phonetically very similar, if not identical, when they appear in isolation (Vance 1995; Warner 1997); in the case of disyllabic words, for example, both finally-accented words and unaccented words receive a LH contour. This is why when examples are shown, a nominative particle suffix [+ga] is often attached: by providing

an extra syllable at the end, it allows us to make clear the distinction between final accented words and unaccented words. (Not all particles are tonally-neutral, however; see section 6).

Unlike some other tonal languages, which can have tonal contrasts on all syllables, Japanese allows only one HL pitch fall within a word; this restriction – that there can maximally be one prominence within a word – is sometimes called "culminativity" (Alderete 1999b; Hayes 1995; Hyman 2009; Ito and Mester 2003, 2012; Revithiadou 1999; Trubetzkoy 1939/1969 – see also Ishihara, this volume). In the context of Japanese, the culminativity restriction means that there can be at most one accentual HL fall.[8] Given this culminativity restriction, the whole tonal contour of words can be predicted as long as the location of accent (and the presence thereof) is known. This limited use of tonal contours is a primary reason for considering Japanese a pitch accent language rather than a tonal language (but see Hyman 2009 for arguments against this view; see also Hulst 2011 for further discussion on this debate).

Since, as illustrated in this section, the tonal contour of a word can be determined based on its accentual properties, the rest of this chapter provides accentual representations only.

2 Loanword accentuation: a default accent pattern

Although the distribution of Japanese accent is often considered to be unpredictable, as the examples in (1) and (2) show, there are environments in which the presence and the location of accent are more or less predictable. This chapter focuses on such predictable patterning. We will start with loanword patterns in this section, which arguably instantiate a default accentuation pattern in Japanese (see Kubozono, Ch. 8, this volume, for more on loanword accent). Here, the studies on loanword accentuation are making a general, but not uncontroversial, assumption that loanword adaptation is a natural, real-world "wug-test" (Berko 1958), in which speakers are forced to pronounce words that they have not encountered before (Kang 2011). Wug-tests are known to be a good tool to reveal speakers' grammatical knowledge (see Kawahara 2011 for a recent overview).

8 One exception is phrasal compounds which allow more than one accent. Many such examples are right branching compounds with three elements (e.g., /[si'n+[nihon+pu'roresu]]/ 'New Japan Wrestling') (Ito and Mester 2007; Kubozono, Ito, and Mester 1997). These compounds arguably involve more than one Prosodic Word (Ito and Mester 2007), which suggests that culminativity should be perhaps determined over a phonological Prosodic Word (or a Minor Phrase), rather than a morphological word. See also Ishihara (this volume) for further discussion on culminativity in Japanese.

2.1 Basic patterns

Loanwords are vocabulary items that Japanese speakers have recently borrowed from other languages, mainly from English (see Kubozono, Ch. 8, this volume; see Kang 2011 for more general discussion on loanword adaptation and loanword phonology). When new words are borrowed into Japanese, they do not have lexical specification for accent. Therefore, Japanese speakers were/are free to assign an accent pattern at their disposal.[9] For this reason, loanword accent provides a window into the default accentuation pattern in Japanese. As a first noticeable characteristic of loanwords, they are more frequently accented than native words; according to Kubozono (2008), 93% of the loanwords in his corpus (N = 778) are accented, whereas only 29% of the native words (N = 2,220) are accented. Kubozono (2006) hypothesizes that when Japanese speakers borrow English words, they hear English pitch patterns in citation forms and map that percept of prominence as Japanese accent, but that its location is determined by the phonological grammar of Japanese.

The locations of accent in loanwords are more or less predictable. Some typical examples are shown in (10), and they are all accented on the antepenultimate mora (the third from the end), which is shown in bold (recall that mora boundaries are shown by -).

(10) Accent assigned on the antepenultimate moras in loanwords
a. ku-ri-**su’**-ma-su ‘Christmas’
b. a-pa-**ra’**-ti-a ‘Appalachia’
c. a-n-da-**ru’**-si-a ‘Andalusia’
d. o-o-su-to-**ra’**-ri-a ‘Australia’
e. o-o-su-**to’**-ri-a ‘Austria’
f. su-**to’**-re-su ‘stress’
g. a-su-**fa’**-ru-to ‘asphalt’
h. ma-ku-do-**na’**-ru-do ‘McDonald’
i. pu-ro-**gu’**-ra-mu ‘program’
j. a-su-pa-**ra’**-ga-su ‘asparagus’
k. pu-ra-**mo’**-de-ru ‘plastic model’
l. e-me-**ra’**-ru-do ‘emerald’
m. zya-a-na-**ri’**-zu-mu ‘journalism’
n. yo-o-**gu’**-ru-to ‘yogurt’
o. a-bu-ra-ka-**da’**-bu-ra ‘Abracadabra’

9 There may be some cases in which Japanese speakers assign accent by mimicking the original English stress pattern. See note 29 for some potential examples. This borrowing pattern can be formally modeled as a faithfulness effect between source forms and borrowed forms (Smith 2007).

In Japanese, any vowel, a coda nasal, and the second half of a geminate are moraic (see Kubozono's introduction to this volume; Kawahara, this volume; Kawagoe, this volume; and Otake, this volume). In (10), accent falls on the antepenultimate mora in the words. This accent pattern is recurrently observed in many loanwords, for which there are arguably no underlying accentual specifications. Therefore, this antepenultimate accent rule has been considered a default accent assignment rule in Japanese (McCawley 1968). For bimoraic forms, there are no antepenultimate syllables, so the accent falls on the penultimate – the second-to-last – syllable (e.g., /mo'ka/ 'mocha' and /mo'ma/ 'MoMA (the Museum of Modern Art)').

From the perspective of modern prosodic phonology (Liberman and Prince 1977; Selkirk 1980 *et seq.*), this antepenultimate accent pattern can be derived by positing a bimoraic trochaic foot (Poser 1990), with the word-final syllable being unfooted; e.g., /kuri(su'ma)su/ (Ito and Mester 1992/2003, 2012; Kawahara and Wolf 2010). See Ito and Mester (2012) and Katayama (1998) for an alternative analysis.

2.2 Syllables as accent-bearing units

When the antepenultimate mora is a so-called deficient (or non-head) mora – the second part of a diphthong (see Kubozono, Ch. 5, this volume), the second half of a geminate or a long vowel, or a coda nasal – the accent does not fall on that mora, and instead shifts to the pre-antepenultimate mora, as the examples in (11) show, in which the antepenultimate moras are shown in bold. A deficient mora combines with the preceding mora and constitutes the second half of a syllable; or differently put, deficient moras are those that do not occupy the head position of a syllable. Based on this observation, McCawley (1968) proposed that the default accentuation in Japanese is that the syllable containing the antepenultimate mora receives accent. For example, /painappuru/ is syllabified as /pai.nap.pu.ru/, and the accent falls on the syllable containing the antepenultimate mora (i.e., /nap/).

(11) Accent assigned on the pre-antepenultimate mora in loanwords
- a. pa-i-na'-**p**-pu-ru 'pineapple'
- b. ta'-**k**-ku-ru 'tackle'
- c. gu-ra'-**n**-pu-ri 'Grand prix'
- d. ka'-**n**-za-su 'Kansas'
- e. ka-re'-**n**-da-a 'calendar'
- f. pu-ri'-**n**-se-su 'princess'
- g. syu-no'-**o**-ke-ru 'snorkel'
- h. pa'-**a**-pu-ru 'purple'
- i. ra'-**i**-fu-ru 'rifle'
- j. ta-i-pu-ra'-**i**-ta-a 'typewriter'
- k. ri-sa'-**i**-ku-ru 'recycle'
- l. bu-ro'-**i**-ra-a 'broiler'

Since we can unify the case in (10) and the case in (11) from a syllable perspective (the accent falls on the syllable containing the antepenultimate mora), the data in (11) support the hypothesis that the bearer of accent is a syllable rather than a mora (McCawley 1968, 1977 – see Labrune 2012 for an alternative view).

Another piece of evidence that syllables bear accent in Japanese comes from the behavior of pre-accenting morphemes, which we will discuss more extensively in Section 6. For example, the suffix /+ke/ 'family of' puts accent on the final vowel of the root to which it is attached, as in (12b–d). When the root-final mora is a non-head of a syllable, however, the accent falls on the penultimate vowel of the root, i.e. the head of the root-final syllable, as in (12e–g). This patterning again shows that syllables bear Japanese accent, not moras.

(12) A dominant pre-accenting suffix inserts accent on the syllable immediately preceding the affix

a.	/+'ke/ 'family of'			
b.	ono	→	ono'+ke	'family of Ono'
c.	yosida	→	yosida'+ke	'family of Yoshida'
d.	edogawa	→	edogawa'+ke	'family of Edogawa'
e.	ku'dan	→	kuda'n+ke	'family of Kudan'
f.	ka'too	→	kato'o+ke	'family of Kato'
g.	ka'sai	→	kasa'i+ke	'family of Kasai'

2.3 The Latin Stress Rule as an alternative formulation?

While the antepenultimate rule explains a good portion of the accentuation patterns in Japanese loanwords, an alternative way to characterize the default accent pattern has been developed in a series of works by Kubozono and others (Haraguchi 1991, 1999; Kubozono 1996, 1999, 2008, 2011; Shinohara 2000; see also Kubozono, Ch. 8, this volume). These works capitalize on the similarity between the antepenultimate accent rule and the Latin Stress Rule (Hayes 1995; Mester 1994). The Latin Stress Rule, which is arguably operative in many languages (Hayes 1995), states that the penultimate syllable is stressed if heavy, but that the antepenultimate syllable is stressed otherwise. Crucial to this rule is the notion of syllable weight – setting aside cross-linguistic complications (Gordon 2002; Hayes 1989, 1995; Rosenthall and van der Hulst 1999; Zec 1995), in Japanese, syllables containing a coda consonant (a moraic nasal or the first part of geminate), a long vowel or a diphthong are bimoraic and heavy, whereas open syllables with short vowels are monomoraic and light. For example, /tan, tat, taa, too, tai, toi/ are all heavy, whereas /ta/ is light.

We can now compare the antepenultimate accent rule (AAR) and the Latin Stress Rule (LSR). Let H represent heavy syllables and L light syllables. Table 1 compares the predictions of these two rules for trisyllabic words with all possible syllable

weight compositions. We observe that in six out of eight conditions, these two rules make the same predictions. Only in two conditions (HLH and LLH) do the two theories make different predictions.

Table 1: Comparing the predictions of the antepenultimate accent rule (AAR) and Latin Stress Rule (LSR). H=Heavy syllable; L=Light syllable.

	AAR	LSR	(mis)match
a.	HH'H	HH'H	match
b.	HH'L	HH'L	match
c.	HL'H	H'LH	mismatch
d.	H'LL	H'LL	match
e.	LH'H	LH'H	match
f.	LH'L	LH'L	match
g.	LL'H	L'LH	mismatch
h.	L'LL	L'LL	match

Kubozono (1996, 1999, 2008, 2011) points out that even in these two mismatching conditions, the forms that are predicted by LSR are actually observed. Some of these LSR-conforming forms appear as variants of the pronunciations predicted by AAR, as the examples in (13) and (14) show, although there are forms that are predicted only by AAR too, as in (15). Katayama (1998) and Kubozono (2008) further argue that the forms (or renditions) that conform to LSR are more common than those that follow AAR, suggesting that the default accentuation pattern in Japanese could be the Latin Stress Rule.[10]

(13) HLH words whose accent locations are predicted by LSR

a. be'e.ka.rii 'bakery'
b. ma'a.ga.rin 'margarine'
c. po'o.to.ree 'portray'
d. my'uu.zi.syan~myuu.zi'.syan 'musician'
e. ha'n.ga.rii~han.ga'.rii 'Hungary'
f. e'n.de.baa~en.de'.baa 'endeavor'
g. o'o.di.syon~oo.di'.syon 'audition'
h. ka'a.de.gan~kaa.de'.gan 'cardigan'
i. ra'n.de.buu~ran.de'.buu 'rendez-vous'
j. ba'n.ga.roo~ban.ga'.roo 'bungalow'
k. pyu'u.ri.tan~pyuu.ri'.tan 'Puritan'

10 Two caveats: (i) LSR does not allow for unaccented outcomes, while Japanese does (see section 2.4 and Ito and Mester 2012); (ii) when words with a sequence of four light syllables (LLLL) are accented, the accent can fall on the pre-antepenultimate mora, as in /bi'zinesu/ 'business' and /a'kusesu/ 'access'. The pre-antepenultimate pattern in this type of word is not predicted by LSR (or by AAR either). It is possible that the final vowels of these words may be invisible to the accent assignment rule since they tend to be epenthetic (Kubozono 1996, 2001b).

(14) LLH words whose accent locations are predicted by LSR

a.	do’.ku.taa	‘doctor’
b.	ma’.su.taa	‘master’
c.	pi’.re.nee	‘the Pyrenees’
d.	te’.he.ran	‘Teheran’
e.	te’.ne.sii	‘Tennessee’
f.	a’.ma.zon	‘Amazon’
g.	me’.ru.hen	‘Fairly tale’
h.	to’.ro.fii	‘trophy’
i.	su’.ri.raa~su.ri’.raa	‘thriller’
j.	do’.ra.gon~do.ra’.gon	‘dragon’
k.	re’.ba.non~re.ba’.non	‘Lebanon’
l.	ma’.zi.syan~ma.zi’.syan	‘magician’
m.	e.ne’.ru.gii~e.ne.ru’.gii	‘energy’

(15) LLH and HLH forms that follow AAR

a.	bi.ta’.min	‘vitamin’
b.	a.se’.an	‘ASEAN (Association of SouthEast Asian Nations)’
c.	hi.ro’.in	‘heroin’
d.	bu.re’.zaa	‘brazier’
e.	su.pu’.ree	‘spray’
f.	bu.ra’.zyaa	‘bra (brassiere)’
g.	baa.be’.kyuu	‘barbecue’
h.	kuu.de’.taa	‘coup’
i.	kon.di’.syon	‘condition’

If the Japanese default accentuation rule is indeed the LSR, then Japanese is a weight-sensitive language in which heavy syllables attract metrical prominence. This cross-linguistically widely observed pattern – the requirement that heavy syllables receive metrical prominence – is called the Weight-to-Stress Principle (WSP) (Hayes 1995; Prince 1983, 1990; Prince and Smolensky 1993/2004). Furthermore, this weight-sensitivity may explain why loanwords are much more likely to be accented than native words (Ito and Mester 2012; Kubozono 1996, 2006, 2008; Sibata 1994). Kubozono (2008) argues that loanwords contain many more heavy syllables than native words (see Nasu, this volume, and Kubozono, Ch. 8, this volume), and that because of the WSP, there are many more accented loanwords.

2.4 Unaccented loanwords

Although loanwords are generally pronounced with accent, as we observed in the previous discussion, there are particular phonological environments in which unaccented words appear. One is the case of four-mora words with two final light syllables, where both of the last two vowels are non-epenthetic, as shown in (16) (Kubozono 1996, 2010, 2011; Kubozono and Ogawa 2005, see also Ito and Mester

2012). This pattern should be contrasted with the cases in (17), where either or both of the final two vowels are epenthetic (shown by < >), and (18), where either of the last two syllables is a heavy syllable.[11]

(16) Unaccented loanwords: four mora words with two final light non-epenthetic syllables

a. a.me.ri.ka 'America'
b. i.ta.ri.a 'Italia'
c. me.ki.si.ko 'Mexico'
d. ai.o.wa 'Iowa'
e. a.ri.zo.na 'Arizona'
f. ai.da.ho 'Idaho'
g. mo.su.ku.wa 'Moscow'
h. ma.ka.ro.ni 'macaroni'
i. kon.so.me 'consommé'
j. mo.na.ri.za 'Mona Lisa'
k. an.te.na 'antenna'

(17) The presence of an epenthetic vowel results in accented words

a. a'n.de.s<u> 'Andes'
b. u.we'.r<u>.z<u> 'Wales'
c. si'n.ba.r<u> 'cymbal'
d. si'n.bo.r<u> 'symbol'
e. a'i.do.r<u> 'idol'
f. p<u>.ro'.se.s<u> 'process'
g. he'e.ge.r<u> 'Hegel'
h. ma'.r<u>.k<u>.s<u> 'Marx'

(18) Penultimate or final heavy syllables result in accented words

a. pa.re'e.do 'parade'
b. o.re'n.zi 'orange'
c. go.bi'n.da 'Govinda (personal name)'
d. o.ha'i.o 'Ohio'
e. i.ra'i.za 'Eliza'
f. e.ri'i.ze 'Elise'
g. ro'n.don 'London'
h. su.to'.roo 'straw'
i. bi.ta'.min 'vitamin'
j. a.se'.an 'ASEAN'

11 Given LHL words, if the first vowel is epenthetic and the final vowel syllable is /to/ or /do/ with epenthetic /<o>/, they can often be unaccented; e.g., /s<u>keet<o>/ 'skate', /p<u>reet<o>/ 'plate' and /p<u>raid<o>/ (Kubozono and Ohta 1998).

There are a number of exceptions to these generalizations as well, however. The words in (19) are pronounced as unaccented, despite the fact that the final vowels are all epenthetic. The words in (20) are also unaccented, despite the fact that their penultimate syllables are heavy.

(19) Unaccented nouns with epenthetic vowels
  a. bu.ra.zi.r<u> ‘Brazil’
  b. boo.ka.r<u> ‘vocal’
  c. san.da.r<u> ‘sandal’
  d. ka.ta.ro.g<u> ‘catalog’
  e. o.mu.re.t<u> ‘omelet’

(20) Unaccented nouns with heavy syllables
  a. hu.ran.su ‘France’
  b. o.ran.da ‘Holland’
  c. ku.ree.mu ‘claim’
  d. hu.ree.zu ‘phrase’

Setting aside these complications, the emergence of unaccented forms in four-mora words is observed in compounds created by truncation as well, as we will observe in section 7.2 (see also Kubozono, Ch. 8, this volume, for more data about unaccented words). See Ito and Mester (2012) and Tanaka (2001) for analyses of the emergence of unaccented patterns in Japanese.

Another case in which unaccented renditions of words appear is so-called *senmonka akusento* (or ‘specialists’ accent’), in which common jargon terms within a certain community tend to be pronounced as unaccented, even when they are pronounced as accented outside of that community (Akinaga 1985; Inoue 1998; Labrune 2012). For example, two loanwords, /rake’tto/ ‘racket’ and /sa’abisu/ ‘service’, are usually accented, but those who engage in playing tennis can and often do pronounce these words as unaccented. Likewise, many computer jargon terms like /sukuriputo/ ‘script’, /purintaa/ ‘printer’ and /puroguramu/ ‘program’ are often pronounced as unaccented. *Zuuzya-go* (or *zuuja-go*), a secret language among musicians (Ito, Kitagawa, and Mester 1996), often results in unaccented words, which again may be an instance of *senmonka akusento*. Finally, phonologists can talk about “autosegmental spreading” as /supuredingu/ and ‘constraint ranking’ as /rankingu/, both pronounced as unaccented.

This *senmonka akusento* resulted in some minimal pairs in terms of the presence of accent in loanwords. For example, /pa’ntu/ means ‘underwear’, whereas /pantu/ (unaccented) can mean ‘trousers’ (in the field of fashion). Similarly, /ku’rabu/ means ‘groups in extracurricular activities (in schools)’ whereas /kurabu/ means ‘(night) club’, and /sa’akuru/ means ‘circle’ but /saakuru/ means ‘extracurricular groups (in colleges)’.

3 Stochastic skews in native and Sino-Japanese nouns

The general assumption about Japanese accent, at least for native and Sino-Japanese (SJ) nouns,[12] is that its distribution is not predictable, as there are examples like (1) and (2) (e.g. /ka'ki/ 'oyster' vs. /kaki'/ 'fence' vs. /kaki/ 'persimmon'), although there are some regularities concerning the accent distributions in loanwords. Kubozono (2006, 2008, 2011) challenges this view, pointing out that there is a stochastic skew in the Japanese lexicon already, which hints at the antepenultimate accent pattern. Many native nouns and SJ nouns are actually unaccented: 71% of native nouns (N = 2,220) and 51% of Sino-Japanese nouns (N = 4,939) in his database. If we look at only accented nouns and examine the distribution of accent locations, an interesting pattern emerges. Consider Table 2, which is adapted from Kubozono (2008: 170).

Table 2: Distributions of different accent patterns in trisyllabic words

Accent pattern	antepenultimate	penultimate	final	N
Native	59%	33%	9%	634
SJ	95%	2%	3%	2,427
Loanwords	96%	2%	2%	722

We observe that in Sino-Japanese nouns, antepenultimate accent is the dominant pattern. Even in native words, more than half of the accented nouns have the antepenultimate accent. In both cases, the accent patterns in the Japanese lexicon are skewed toward antepenultimate accent. These observations show that the default accentuation assignment rule in loanwords may not have come out of the blue, but came instead from an abstraction over the distributional skew that already existed in the lexicon at the time of loanword adaptation.

For a more comprehensive analysis of distributional skews of accent types for words with different lengths, see Sibata (1994), translated into English by Labrune (2012), as well as Kitahara (2001), further analyzed by Ito and Mester (2012).

4 Compound accent

Compound accent is arguably one of the most extensively discussed areas of research in Japanese accentology. A traditional view of this research categorizes compound accent rules into two cases according to the phonological length of

12 Sino-Japanese nouns are borrowings from old Chinese words. See Ito and Mester (1995, 1996, 1999, 2008) as well as Kawagoe (this volume) and Ito and Mester (Ch. 7, this volume).

second elements (N2) (Akinaga 1985; McCawley 1968; Poser 1990), where a short N2 is either monomoraic or bimoraic. Although there have been attempts to unify these cases (Kubozono 1995, 1997, 2008; Kubozono and Mester 1995), the discussion here follows this traditional dichotomy.

4.1 Short N2

Short nouns are either monomoraic or bimoraic. They can behave in two ways:[13] those that retain their accent, or those that assign accent on the last syllable of the N1, as exemplified in (21) and (22), respectively.[14] Labrune (2012) and Tanaka (2001) provide more examples of each type of these N2s.

(21) Short N2 that retain their N2 accent[15]

a.	fa’asuto+ki’su	→	faasuto+ki’su	‘first kiss’
b.	koosoku+ba’su	→	koosoku+ba’su	‘Highway bus’
c.	tennen+ga’su	→	tennen+ga’su	‘natural gas’
d.	kyooiku+ma’ma	→	kyooiku+ma’ma	‘education-minded mama’
e.	ku’kkingu+pa’pa	→	kukkingu+pa’pa	‘cooking papa’
f.	niho’n+ha’mu	→	nihon+ha’mu	‘Japan ham’
g.	boohan+be’ru	→	boohan+be’ru	‘security alarm’
h.	niho’n+sa’ru	→	nihon+za’ru	‘Japan monkey’
i.	takara’+hu’ne	→	takara+bu’ne	‘treasure ship’
j.	pe’rusya+ne’ko	→	perusya+ne’ko	‘Persian cat’
k.	garasu+ma’do	→	garasu+ma’do	‘glass window’

(22) Pre-accenting short N2

a.	ka’buto+musi	→	kabuto’+musi	‘beetle’
b.	minasi+ko	→	minasi’+go	‘orphan’
c.	ma’igo+inu’	→	maigo’+inu	‘lost puppy’
d.	undoo+kutu’	→	undo’o+gutu	‘exercise shoes’
e.	kana’gawa+si’	→	kanagawa’+si	‘Kanagawa City’
f.	sa’rada+ba’a	→	sarada’+baa	‘Salad bar’
g.	kuri’imu+pa’n	→	kuriimu’+pan	‘custard bread’
h.	hirosima+ke’n	→	hirosima’+ken	‘Hiroshima Prefecture’
i.	ni’ngyo+hi’me	→	ningyo’+hime	‘Little Mermaid’

13 For now we set aside deaccenting morphemes, and will come back to them in section 6.

14 When a compound consists of bimoraic N1 and bimoraic N2, resulting in compounds with 4 moras, we often observe an unaccented outcome: /neko+basu/ ‘cat bus’. See section 7.2 and Kubozono and Fujiura (2004).

15 In some compound forms, the first consonant of N2 becomes voiced. This phenomenon is called “rendaku”. See Vance (this volume) for extensive discussion of this phenomenon.

j.	sizyu'u+ka'ta	→	sizyu'u+kata	'forty(-year-old)'s shoulder (adhesive capsulitis)'
k.	na'iron+i'to	→	nairo'n+ito	'nylon thread'
l.	niwaka+a'me	→	niwaka'+ame	'sudden rain'
m.	yoyaku+se'ki	→	yoyaku'+seki	'reserved seats'
n.	ueno+e'ki	→	ueno'+eki	'Ueno station'
o.	gakusyuu+zyu'ku	→	gakusyu'u+zyuku	'learning prep-school'

All the examples in (21), which retain their N2 accent, have accent on their penultimate syllables, whereas many of the N2s in (22) are unaccented or have final accent (=(22a–h)). Kubozono (1995, 1997, 2008), building on Poser (1990), points out that when N2 bears accent on its final syllable, it very often loses its accent and becomes pre-accenting.[16] Kubozono (1995, 1997) attributes this loss of final accent to a constraint against having prominence at word edges (also known as NonFinality(σ): Prince and Smolensky 1993/2004; Hyde 2007, 2011).

For N2s which have non-final accent, there is lexical variation: those that retain their accent like /ne'ko/, as in (21), and those that lose their accent, like /hi'me/, as in (22i–o). Furthermore, the last two forms (=(21j–k)) may allow the pre-accenting pronunciation as a variant form. The fact that some items lose their penultimate accent indicates that penultimate accent, which is in the final foot, are marked.[17] This effect can be attributed to another sort of NonFinality constraint: i.e. NonFinality(Ft) (Kawahara and Wolf 2010; Kubozono 1995, 1997; Kurisu 2005; Shinohara 2000). The remaining issue is how to model the item-specific behavior in terms of whether they are allowed to violate NonFinality(Ft) (=(21)) or not (=(22)), which is a general challenge to phonological theory (Coetzee 2009; Inkelas 1999; Inkelas and Zoll 2007; Inkelas, Orgun, and Zoll 1997; Kisseberth 1970; Pater 2000, 2010, among many others).

Among those that retain N2 accent, many of the examples are of foreign origin (i.e., loanwords) (see Tanaka 2001 for details). In (21), more than half of the examples involve a loanword N2 (=21a–g). The retention of N2 accent may thus partly be due to a faithfulness effect specific to loanwords (Ito and Mester 1999, 2008). The fact that few if any loanwords lose their penultimate accent – no words in (22) are loanwords – supports this idea (see Kubozono, Ch. 8, this volume, for additional evidence). Finally, Sino-Japanese words, (22m–o), almost always lose their N2 penultimate accent (Kawahara, Nishimura, and Ono 2002; Kubozono 1997; Tanaka 2002). To summarize, there are differences among different lexical classes in

16 There are exceptions, which retain the final accent of N2; e.g., *kenkyuu+zyo'* 'research center', *keisatu+syo'* 'police station' and *bitamin-si'i* 'Vitamin C' (Tanaka 2001).

17 Regardless of whether the final syllable is footed (e.g., *ningyo-(hi'me)*) or not (e.g., *nin(gyo-hi')me*), the penultimate accent is in the final foot.

terms of the likelihood of the attrition of N2 penultimate accent: Sino-Japanese > native words > loanwords.[18]

4.2 Long N2

When N2 is trimoraic or longer, there are two major generalizations: (i) if N2 is unaccented or has accent on the final syllable, then the accent falls on the initial syllable of N2, as in (23); (ii) otherwise, the accent of N2 is retained, as in (24).

(23) N2 initial accent

a.	si'n+yokohama	→	sin+yo'kohama	'Shin-Yokohama (place name)'
b.	minami+amerika	→	minami+a'merika	'South America'
c.	ko'o+ketuatu	→	koo+ke'tuatu	'high blood pressure'
d.	onna'+tomodati	→	onna+to'modati	'female friend'
e.	kuti+yakusoku	→	kuti+ya'kusoku	'verbal promise'
f.	dame'+otoko'	→	dame+o'toko	'unreliable men'
g.	de'ka+atama'	→	deka+a'tama	'big head'
h.	nise+takara'	→	nise+da'kara	'fake treasure'

(24) Retention of N2 accent

a.	si'n+tamane'gi	→	sin+tamane'gi	'new onion'
b.	ya'mato+nade'siko	→	yamato+nade'siko	'Japanese lady'
c.	be'suto+hure'ndo	→	besuto+hure'ndo	'best friend'
d.	a'ka+ore'nzi	→	aka+ore'nzi	'red-orange'
e.	tuukin+sarari'iman	→	tuukin+sarari'iman	'commuting salaryman'
f.	natu'+kuda'mono	→	natu+kuda'mono	'summer fruits'

Moreover, for those N2 that have penultimate accent, there can be some variation (Kubozono 2008), as exemplified in (25). As is the case for short N2, it seems that penultimate accent in N2 in compounding may be marked (i.e., the effect of NonFinality(Ft)).

(25) Variation between initial accenting and retention of N2 accent

a.	na'ma+tama'go	→	nama+ta'mago~nama+tama'go	'raw egg'
b.	kami'+omu'tu	→	kami+o'mutu~kami+omu'tu	'paper diaper'
c.	hidari+uti'wa	→	hidari+u'tiwa~hidari+uti'wa	'being luxurious'

18 This observation counter-exemplifies the proposal by Ito and Mester (1999) that faithfulness constraints for Sino-Japanese are always ranked above faithfulness constraints for native words (Kawahara, Nishimura, and Ono 2002).

Given what we have seen for short N2 and long N2, some general tendencies emerge. First, as in (22) and (23), accent on the final syllables of N2 tends to get lost, and new compound accent is assigned (except for some exceptions noted in note 16). As stated above, this pressure is perhaps a reflection of a cross-linguistic tendency to avoid final prominence. Since final accent is allowed in free-standing lexical items, Japanese compound accentuation is a case of "the emergence of the unmarked" (Becker and Flack 2011; McCarthy and Prince 1994) in morphologically derived environments, in which only unmarked structures are allowed in particular (phonological or morphological) environments.

Accent on the final foot is avoided but can be tolerated, as shown by the difference between (21) and (22) as well as the variability in (25). For example, in *ningyo+(hi'me)*, the accent in the final foot is marked, and therefore a new compound accent is assigned for the forms in (22). Accent on final syllables is more likely to be avoided than accent on final feet, which indicates that NONFINALITY(σ) and NONFINALITY(FT) are separate constraints (Kubozono 1995, 1997; Tanaka 2001).

Finally, to complete the picture, when N2 is longer than 4 moras, the compound accent tends to simply retain the accent of N2 (Kubozono, Ito, and Mester 1997; Labrune 2012). Even when N2 is unaccented, it does not result in N2-initial accent, unlike the forms in (23). This avoidance of N2-initial accent may be related to a ban on putting accent on a syllable that is too far away from the right edge of a word.

(26) Superlong N2

a.	si'donii+orinpi'kku	→	sidonii+orinpi'kku	'Sydney Olympics'
b.	iso'ppu+monoga'tari	→	isoppu+monoga'tari	'Aesop's Fables'
c.	minami+kariforunia	→	minami+kariforunia	'Southern California'
d.	nyu'u+karedonia	→	nyuu+karedonia	'New Caledonia'
e.	nankyoku+tankentai	→	nankyoku+tankentai	'South Pole expedition team'

There have been extensive theoretical analyses of compound accent patterns – and other related accentual phenomena – in Japanese, especially within the framework of Optimality Theory (Prince and Smolensky 1993/2004), building on the patterns reviewed in this section. Readers are referred to this body of literature for further details (Ito and Mester 2003, 2007, 2012; Kawahara and Wolf 2010; Kubozono 1995, 1997, 2008, 2011; Kubozono, Ito, and Mester 1997; Labrune 2012; Poser 1990; Shinohara 2000; Tanaka 2001).

5 Verbs and adjectives

Compared to the accent patterns of nouns, the accentual properties of verbs and adjectives are relatively simple. Concretely, verbs and adjectives do not contrast in

terms of the location of accent; rather, the contrast is simply a matter of accented vs. unaccented. The examples in (27) and (28) illustrate this contrast, using the non-past forms for illustration. In recent years, unaccented adjectives are becoming accented, especially among young speakers, which results in the neutralization of the accentual contrast in adjectives (Akinaga 2002; Kobayashi 2003).[19]

(27) Verb accent

a.	moe'+ru	'to come into blossom'	vs	moe+ru	'to fire'
b.	ki'r+u	'to cut'	vs.	ki+ru	'to wear'
c.	na'r+u	'to become'	vs.	nar+u	'to ring'
d.	hare'+ru	'to be sunny'	vs.	hare+ru	'to be swollen'
e.	yoroko'b+u	'to be pleased'	vs.	utaga+u	'to doubt'

(28) Adjective accent

a.	atu'+i	'hot'	vs.	atu+i	'thick'
b.	uma'+i	'delicious'	vs.	ama+i	'sweet'
c.	tanosi'+i	fun'	vs.	tumeta+i	'cold'
d.	omosiro'+i	'funny'	vs.	usugura+i	'slightly dark'

As observed in (27) and (28), the location of accent for accented words is on the penultimate mora.[20] Since verbs and adjectives inevitably come with inflectional endings in Japanese, one could imagine that some mechanism similar to the compound accent rule for short N2 is operative, as in (22) (Kubozono 2008).

However, when we consider a full set of inflected forms, the story becomes more complicated. Japanese regular verbs are classified into two sets, V-final roots and C-final roots, and they behave slightly differently in terms of accentuation (see also Ito and Mester, Ch. 9, this volume). First, Table 3 shows example inflectional paradigms for V-final verbs. In the case of an accented verb, the accent falls on the penultimate, root-final syllable in the negative and conditional forms. These suffixes may be accent-shifting suffixes. In the polite and volitional forms, the accent shifts to the suffix. Such a suffix is called a dominant suffix; see section 6 for more on these types of suffixes. Another interesting puzzle is that the accent shifts to the antepenultimate position in the gerundive and past forms (McCawley 1968; Yamaguchi 2010).

19 Based on a sociolinguistic production study, Kobayashi (2003) found that among other factors, sonority of the penultimate syllable affects this sound change in such a way that the less sonorous the consonant in the penultimate syllable is, the more likely it is that the word becomes accented. This pattern is parallel to onset-driven stress patterns where syllables with low-sonority onsets attract stress (e.g., Gordon 2005 and references cited therein).

20 Similar to the case of loanword accentuation, when the penultimate mora is a second part of a syllable, the accent shifts one mora leftward to the antepenultimate mora; e.g., /ha'i+ru/ 'enter' and /to'o+ru/ 'go through' (Vance 1987). However, there is a (near) minimal pair like /ka'e+ru/ 'to return' and /hae'+ru/ 'to reflect', which adds another layer of complication (Yamaguchi 2010).

In the case of unaccented V-final roots, most forms are unaccented except when one of the two dominant suffixes is attached (the polite form and the volitional form). In addition, the conditional suffix /+reba/ shows its accent only when it is attached to unaccented roots – this suffix is a recessive suffix. Again see section 6.

Table 3: Verb inflection table: V-final roots

	Accented 'to be sunny'	Unaccented 'to be swollen'
negative	hare'+nai	hare+nai
polite	hare+ma'su	hare+ma'su
non-past	hare'+ru	hare+ru
gerundive	ha're+te	hare+te
past	ha're+ta	hare+ta
conditional	hare'+reba	hare+re'ba
volitional	hare+yo'o	hare+yo'o

Consonant-final verbs behave slightly differently, as shown in Table 4. For accented roots, the polite, non-past, conditional, and volitional forms pattern the same as with V-final roots. Two differences are (i) in the negative form, the accent falls on the suffix-initial vowel, and (ii) in the gerundive and past tense forms, no shift to the antepenultimate position occurs. C-final unaccented roots behave much the same way as V-final unaccented roots.

Table 4: Verb inflection table: C-final roots

	Accented 'to be pleased'	Unaccented 'to work'
negative	yorokob+a'nai	hatarak+anai
polite	yorokob+ima'su	hatarak+ima'su
non-past	yoroko'b+u	hatarak+u
gerundive	yoroko'n+de	hatarai+te
past	yodoko'n+da	hatarai+ta
conditional	yoroko'b+eba	hatarak+e'ba
volitional	yorokob+o'o	hatarak+o'o

As observed, accent patterns in the various inflectional forms of Japanese verbs are complex. Accordingly, there are a number of analyses of verbal accent patterns (Clark 1986; Haraguchi 1999; McCawley 1968; Nishiyama 2010; Yamaguchi 2010).

Table 5 illustrates typical inflectional paradigms for adjectives.[21] In the inflected forms of an accented adjective, the accent falls on the penultimate mora of the root (not the word). It is not root-initial accentuation, as shown by a longer root,

21 There is non-negligible variation in adjective accent (Akinaga 1985; Martin 1967), which is abstracted away from here, due to space limitations.

like /tano'si-sa/. For unaccented roots, some suffixes (suspensive and conditional) assign accent on the root-final syllable – a case of pre-accentuation. See again section 6.

Table 5: Adjective inflection table

	Accented 'delicious'	Unaccented 'sweet'
non-past	uma'+i	ama+i
deverbal	u'ma+sa	ama+sa
suspensive	u'ma+kute	ama'+kute
adverbial	u'ma+ku	ama+ku
conditional	u'ma+kereba	ama'+kereba

6 Accent patterns of affixes

Several studies, on Japanese and on other languages, have examined how affixes interact with roots in terms of accent. This section introduces various types of affixes that interact with root accent in different ways, as we saw some examples already in section 5. There are many types of affixes in Japanese in terms of their accentual behaviors (Alderete 1999b; Kurisu 2001; McCawley 1968; Poser 1984). The following description draws on Poser (1984) and Vance (1987), and discusses the following eight types of affixes: (i) recessive suffixes, (ii) dominant suffixes, (iii) recessive pre-accenting suffixes, (iv) dominant pre-accenting suffixes, (v) accent shifting suffixes, (vi) post-accenting prefixes, (vii) deaccenting suffixes, and (viii) initial accenting suffixes.

First, we start with the recessive suffix.[22] Recall that Japanese allows one accent per word (culminativity). Therefore, when two morphemes with accent are concatenated, one accent has to be deleted. In such cases, a recessive suffix loses its accent.[23] In other words, it is accented only when it is attached to unaccented roots, as in (29b–d), but it loses its accent when the root is accented, as in (29e–g). This recessive behavior may reflect general tendencies in natural languages to preserve more information from roots than from affixes (Alderete 1999b, 2001b; Beckman 1998; McCarthy and Prince 1995; Urbanczyk 2006, 2011). Another example of this kind of suffix is /+na'do/ 'etc' (Vance 1987).

22 Whether a particular morpheme is an affix or a clitic (or even a bound morpheme root) is controversial, but this chapter sets aside this issue.

23 It may be that the accent deletion results in incomplete neutralization in which some trace of underlying accentedness may be left at the surface (Matsumori et al. 2012: 53–54, see also Igarashi, this volume). For recent reviews of incomplete neutralization, see Braver (2013), Kawahara (2011) and Yu (2011).

(29) A recessive suffix: suffix loses its accent if attached to an accented root

a.	/+ta'ra/ 'conditional'			
b.	her+ta'ra	→	het+ta'ra	'if decreased'
c.	ne+ta'ra	→	ne+ta'ra	'if sleep'
d.	mage+ta'ra	→	mage+ta'ra	'if bent'
e.	tabe'+ta'ra	→	ta'be+tara	'if eat'
f.	nage'+ta'ra	→	na'ge+tara	'if throw'
g.	nagare'+ta'ra	→	naga're+tara	'if flow'

Unlike a recessive suffix, the dominant suffix retains its accent regardless of whether the root is unaccented or not. /+ppo'i/ is an example of this kind – it is accented both when the root is unaccented (30b–d) and when it is accented (30e–g). /+gu'rai/ 'at least' behaves in the same way in that it deletes the root accent to retain its own accent (Vance 1987). In this sense, these suffixes behave like those N2 nouns that retain their accent in compound formation (see (21)). The behavior of these suffixes is different from the general tendency to preserve information from roots, and hence has been analyzed as a result of additional grammatical mechanisms (Alderete 1999b, 2001a; Kurisu 2001).

(30) A dominant suffix: suffix bears accent, and causes deletion of root accent

a.	/+ppo'i/ '-ish'			
b.	abura	→	abura+ppo'i	'oily'
c.	kaze	→	kaze+ppo'i	'sniffly'
d.	kodomo	→	kodomo+ppo'i	'childish'
e.	ada'	→	ada+ppo'i	'coquettish'
f.	netu'	→	netu+ppo'i	'feverish'
g.	ki'za	→	kiza+ppo'i	'snobbish'

The next type of suffix is the pre-accenting suffix, and there are three sub-types: recessive, dominant, and accent-shifting. Pre-accenting suffixes insert accent on the root-final syllable. A recessive suffix of this type, exemplified in (31), inserts accent to its immediately preceding syllable when the root is unaccented as in (31b–d), but does not do so when the root is accented, as in (31e–h).

(31) Recessive pre-accenting: accent inserted on the syllable immediately preceding the suffix, but only if the root is unaccented

a.	/+si/ 'Mr.'			
b.	ono	→	ono'+si	'Mr. Ono'
c.	yosida	→	yosida'+si	'Mr. Yoshida'
d.	edogawa	→	edogawa'+si	'Mr. Edogawa'
e.	u'ra	→	u'ra+si	'Mr. Ura'
f.	mu'raki	→	mu'raki+si	'Mr. Muraki'
g.	nisi'mura	→	nisi'mura+si	'Mr. Nishimura'
h.	tesiga'wara	→	tesiga'wara+si	'Mr. Teshigawara'

The dominant pre-accenting suffix, on the other hand, puts accent on the root-final syllable of both accented and unaccented roots, as in (32). This behavior is similar to those N2 nouns that assign compound accent on the last syllable of N1 (see (22)).

(32) Dominant pre-accenting: accent inserted on the syllable immediately preceding the suffix, regardless of the accent pattern of the root

a.	/+’ke/ ‘family of’			
b.	ono	→	ono’+ke	‘family of Ono’
c.	yosida	→	yosida’+ke	‘family of Yoshida’
d.	edogawa	→	edogawa’+ke	‘family of Edogawa’
e.	u’ra	→	ura’+ke	‘family of Ura’
f.	mu’raki	→	muraki’+ke	‘family of Muraki’
g.	nisi’mura	→	nisimura’+ke	‘family of Nishimura’
h.	tesiga’wara	→	tesigawara’+ke	‘family of Teshigawara’

The third type of pre-accenting suffix inserts accent on the root-final syllable, but only if the root is accented. This suffix does not carry accent of its own, but if a root comes with accent, it attracts that accent immediately to its left. In other words, this suffix can shift already-existing accent, but it cannot insert new accent, unlike other pre-accenting suffixes.

(33) Accent shifting: accent inserted on the syllable immediately preceding the suffix, if the root already has accent

a.	/+mono/ ‘thing’			
b.	ka’k(+u)	→	kaki’+mono	‘thing to write’
c.	yo’m(+u)	→	yomi’+mono	‘thing to read’
d.	tabe’(+ru)	→	tabe’+mono	‘thing to read’
e.	ni(+ru)	→	ni+mono	‘cooked food’
f.	nor(+u)	→	nori+mono	‘thing to ride’
g.	wasure(+ru)	→	wasure+mono	‘forgotten things’

Although Japanese has many more suffixes than prefixes, there are some prefixes, some of which are post-accenting. One example is the honorific prefix /o+/, as in (34) (Haraguchi 1999) (some examples involve truncation of the root materials).[24] Another case of this prefix is /ma+/ (Poser 1984), as exemplified in (35). This suffix causes gemination of the root-initial consonants as well.

24 This post-accentuation has a fair number of exceptions, with /o+/ sometimes behaving as a deaccenting prefix (e.g., *o+ma’nzyuu* → *o+manzyuu* ‘Japanese cake’ and *o+imo’* → *o+imo* ‘potato’), and sometimes behaving as accentually neutral (e.g., *o+misosi’ru* → *o+misosi’ru* ‘miso soup’).

(34) Post-accenting prefix /o+/

a.	/o+/ 'honorific'			
b.	huro'	→	o+hu'ro	'bath'
c.	susi'	→	o+su'si	'sushi'
d.	tegami	→	o+te'gami	'letter'
e.	sentaku	→	o+se'ntaku	'laundry'
f.	kotatu	→	o+ko'ta	'a warm table'
g.	satumaimo	→	o+sa'tu	'potato'
h.	itazura	→	o+i'ta	'trick'
i.	hurui	→	o+hu'ru	'second-handed'
j.	kakimoti	→	o+ka'ki	'rice cracker'

(35) Post-accenting prefix /ma+/

a.	/ma+/ 'truly'			
b.	ma+maru	→	mam+ma'ru	'truly round'
c.	ma+sakasama	→	mas+sa'kasama	'truly downward'
d.	ma+syoome'n	→	mas+syo'omen	'truly face to face'
e.	ma+taira	→	mat+ta'ira	'truly flat'
f.	ma+hiruma	→	map+p'iruma	'noon'
g.	ma+kura'(+i)	→	mak+ku'ra	'truly dark'

There are also morphemes, sometimes called deaccenting morphemes, that result in unaccented words, as in (36).[25] One important generalization about the deaccenting morphemes is that most if not all of them are one or two moras long (e.g., /-iro/ 'color', /-tama/ 'ball', /-too/ '(political) party', etc.) (Akinaga 1985).

(36) Deaccenting: affix bears no accent, and causes deletion of root accent

a.	/+teki/ '-like'			
b.	ke'izai	→	keizai+teki	'economic'
c.	ro'nri	→	ronri+teki	'logical'
d.	go'ori	→	goori+teki	'efficient'
e.	bu'ngaku	→	bungaku+teki	'literature-like'
f.	riki'gaku	→	rikigaku+teki	'in terms of dynamics'
g.	ana'ta	→	anata+teki	'In your opinion (colloquial)'

A local version of this deaccenting behavior is exemplified by the genitive suffix /+no/, which deletes only root-final accent, as in (37d–e) (Haraguchi 1999; Poser 1984). However, there are some complications with this pattern (Vance 1987); for

25 Giriko (2009) points out that there are pseudo-suffixal endings in loanwords that behave as if they are deaccenting suffixes – /(-)in/, /(-)ia/, /(-)ingu/ (e.g., /insurin/ 'insulin', /makedonia/ 'Macedonia', and /ranningu/ 'running').

example, it does not delete accent of a monosyllabic root, as in (37f–h). Further, /+no/ does not tend to delete final accent on heavy syllables, as in (38), although /niho'n/ 'Japan' (38i) is an exception to this sub-generalization.

(37) Local Deaccenting: affix bears no accent, and causes deletion of root-final accent

a.	/+no/ 'GEN'			
b.	i'noti+no	→	i'noti+no	'life+GEN'
c.	koko'ro+no	→	koko'ro+no	'heart+GEN'
d.	atama'+no	→	atama+no	'head+GEN'
e.	kawa'+no	→	kawa+no	'river+GEN'
f.	ha'+no	→	ha'+no	'tooth+GEN'
g.	ki'+no	→	ki'+no	'tree+GEN'
h.	su'+no	→	su'+no	'vinegar+GEN'

(38) Deletion does not target accent of a final heavy syllable

a.	zyapa'n+no	→	zyapa'n+no	'Japan+GEM'
b.	koohi'i+no	→	koohi'i+no	'coffee+GEN'
c.	buru'u+no	→	buru'u+no	'blue+GEN'
d.	wanta'n+no	→	wanta'n+no	'wonton+GEN'
e.	koozyo'o+no	→	koozyo'o+no	'factory+GEN'
f.	hyoozyo'o+no	→	hyoozyo'o+no	'expression+GEN'
g.	masi'n+no	→	masi'n+no	'machine+GEN'
h.	niho'n+no	→	nihon+no	'Japan+GEN'

In addition to these types of suffixes that are recognized in the traditional literature, there may be a new type of suffix, /+zu/, which assigns accent on root-initial syllables, in addition to sometimes lengthening the root-final vowel (Kawahara and Wolf 2010). This suffix is based on a borrowing of the English plural *-s*, and is used to create group names. In some environments at least (Kawahara and Wolf 2010; Kawahara and Kao 2012), this suffix assigns accent on root-initial syllables (see Giriko, Ohshita, and Kubozono 2011 for a reply). This behavior is particularly interesting, since it constitutes a case of non-local interaction between two phonological entities: the suffix and root-initial accent.[26]

26 Some authors claim that cross-linguistically, accent inserted by affixes can land only on adjacent syllables (Alderete 2001a; Kurisu 2001; Revithiadou 2008), but a set of standard assumptions in Optimality Theory (Prince and Smolensky 1993/2004) – in particular, morpheme-specific ALIGNMENT constraints and the existence of ALIGN-L (McCarthy and Prince 1993) – predicts that such a non-local behavior is possible (Kawahara and Wolf 2010; Kawahara and Kao 2012).

(39) Accent pattern of /+zu/

a.	raion	→	ra'ion+zu	'Lions (team name)'
b.	tonneru	→	to'nneru+zu	'Tonneruzu (comedian name)'
c.	okamoto	→	o'kamoto+zu	'Okamotozu (band name)'
d.	heppoko	→	he'ppokoo+zu	'Heppokoozu (personal name)'

Another point worth mentioning is that in the nonce word studies conducted by Kawahara and Kao (2012), initial-accenting was observed more frequently in 4-mora nonce roots (e.g., /husonii+zu/) than 5-mora nonce roots (e.g., /muhusonii+zu/). This difference may be related to the fact that in long nouns, words with initial accent are very rare at best (Kubozono 2008), indicating that Japanese accent is generally right-aligned.

To summarize this section, various types of suffixes interact with root accent in very complex ways. Therefore, modeling the behavior of these different types of suffixes has received some attention in the recent literature (Alderete 1999b, 2001a; Inkelas 2011; Inkelas and Zoll 2007; Kawahara and Wolf 2010; Kurisu 2001; Labrune 2012), especially in the context of Optimality Theory (Prince and Smolensky 1993/2004).

7 Other predictable patterns

This section surveys other domains of Japanese phonology in which accentuation is more or less predictable (see also Akinaga 1985).

7.1 Proper names

Although most proper names – family names and place names – are arguably of native or SJ origin, their accentual properties are more or less predictable, at least more predictable than those of ordinary native nouns (Shinohara 2000). First, names are either accented (40) or unaccented (41), and if accented, the accent falls on the antepenultimate mora; i.e., the default accent location (see section 2). This emergence of the default accentuation in proper names can also be seen in personal names like /sa'kura/ and /hi'nata/, which are accented on the antepenultimate mora, whereas the words that these names are based on are unaccented (/sakura/ 'cherry blossom' and /hinata/ 'sunlight').

(40) Monomorphemic accented names
 a. a’kira
 b. yu’taka
 c. sa’tosi
 d. tu’yosi
 e. ma’doka
 f. a’sina
 g. ta’maki
 h. si’zuka
 i. ho’noka
 j. yosi’masa
 k. take’hiko

(41) Monomorphemic unaccented names
 a. minoru
 b. takeru
 c. manabu
 d. susumu
 e. nagisa
 f. yayoi
 g. sizuku
 h. saori
 i. kaori

In names that are three moras long, those that are derived from adjectives are generally accented (e.g., *tu’yosi* < *tuyo’+i* ‘strong’), whereas those that are derived from verbs (e.g., *minoru* < *mino’r+u* ‘to ripen’) are unaccented (Akinaga 1985).

The accentual properties of first names with a personal suffix are often determined by the suffix. For example, the common female suffix /+ko/ creates accented names, whereas another common female suffix /+mi/ results in unaccented names.[27] If accented, the location is the default – the syllable containing the antepenultimate – as illustrated in (42).

(42) Pairs of accented and unaccented names sharing the same roots

a.	to’mo+ko	vs.	tomo+mi
b.	mi’na+ko	vs.	mina+yo
c.	ha’na+ko	vs.	hana+e
d.	ma’sa+si	vs.	masa+o
e.	si’ge+to	vs.	sige+o
f.	ta’ku+to	vs.	taku+mi
g.	ta’ku+ya	vs.	taku+mi

27 Some suffixes show more complicated behaviors; e.g., /-taroo/ and /-ziroo/ (Kubozono 2001b). Also, /+ko/ shows some irregularity; when it is attached to 3-mora roots, the entire names receive the penultimate accent (e.g., /sakura’+ko/ and /kaoru’+ko/).

7.2 Prosodically truncated words

Japanese exhibits a productive truncation pattern in which long words can be truncated into bimoraic forms, which is arguably a foot-based prosodic morphology pattern (Ito 1990; Ito and Mester 1992/2003; Mester 1990; Poser 1990; see also Ito and Mester, Ch. 9, this volume). This truncation pattern usually, but not always, keeps the first two moras of the original words, and the truncated forms usually have initial accent (Shinohara 2000), whether they are created from native words (43) or loanwords (44). The truncation pattern can truncate personal names into two moras, which results in initially-accented forms (43d–f), as well. See also Mester (1990) and Poser (1990) for other foot-based name forming patterns in Japanese.

(43) Native truncated words (two moras)

a.	na'subi	→	na'su	'eggplant'
b.	tyarinko	→	tya'ri	'bicycle'
c.	moti'ron	→	mo'ti	'of course'
d.	hanae	→	ha'na	'Hanae (personal name)'
e.	ma'sako	→	ma'ko	'Masako (personal name)'
f.	takumi	→	ta'mi	'Takumi (personal name)'

(44) Foreign truncated words (two moras)

a.	demonsutore'esyon	→	de'mo	'demonstration'
b.	tyokore'eto	→	tyo'ko	'chocolate'
c.	riha'asaru	→	ri'ha	'rehearsal'
d.	bi'rudingu	→	bi'ru	'building'
e.	roke'esyon	→	ro'ke	'location'
f.	robo'tto	→	ro'bo	'robot'
g.	terori'zumu	→	te'ro	'terrorism'

In some words, however, truncation keeps the last two moras, in which case the final accent or unaccented outcomes seem to be common, as the examples in (45) show. The last three examples in (45) are all place names, and deaccentuation in (45f–g) may have to do with *senmonka akusento* (section 2.4).

(45) Native truncated words

a.	wa'sabi	→	sabi'	'wasabi'
b.	tomodati	→	dati'	'friend'
c.	syooyu-zuke	→	zuke'	'pickled with soy source'
d.	katura	→	zura(')	'wig'
e.	takara'zuka	→	zuka	'Takarazuka'
f.	sinzyuku	→	zyuku	'Shinjuku (place name)'
g.	takadanoba'ba	→	baba	'Takadanobaba (place name)'
h.	yokohama	→	hama'	'Yokohama (place name)'

When compounds are truncated into two bimoraic feet, the result is usually unaccented. Some examples in (46) and (47) illustrate this pattern.

(46) Native truncated compounds (four moras)

a.	akema'site omedetoo	→	ake+ome	'A Happy New Year'
b.	kotosimo yorosiku	→	koto+yoro	'Keep in touch this year'
c.	tama pura'aza	→	tama+pura	'Tama Plaza (place name)'
d.	hara'(+ga) ita'i	→	hara+ita	'I have a stomachache'
e.	toriatukai setumeesyo	→	tori+setu	'instructions'
f.	tora'nu ta'nuki (no kawaza'n yoo)	→	tora+tanu	'ungrounded profit expectation'

(47) Foreign truncated compounds (four moras)

a.	pa'asonaru kompyu'utaa	→	paso+kon	'personal computer'
b.	mai me'rodii	→	mai+mero	'my melody'
c.	ea kondi'syonaa	→	ea+kon	'air conditioner'
d.	razio kase'tto	→	razi+kase	'radio casette player'
e.	rimo'oto kontoro'oraa	→	rimo+kon	'remote controller'
f.	dezitaru ka'mera	→	dezi+kame	'digital camera'

7.3 Mimetics

Japanese has a large number of sound-symbolic words, which are often referred to as mimetics (see Nasu, this volume). The prosodic shapes and suffixal patterns are regularized in mimetics, and accent patterns are (more or less) predictable for each prosodic pattern (Akinaga 1985; Hamano 1986; Nasu 2002). First, basic forms that appear with /+to/ receive antepenultimate (i.e., the default) accent, as in (48).[28] These roots may appear without the suffixal /+to/, in which case they receive accent on the penultimate mora in the root.

(48) Some mimetic forms

a.	wa't+to	'suddenly'
b.	sa't+to	'swiftly'
c.	kara't+to	'dry'
d.	piri't+to	'stingy'
e.	niko'ri+to	'smily'
f.	hiya'ri+to	'chilly'
g.	ukka'ri+to	'absent-mindedly'
h.	gakka'ri+to	'disappointedly'

28 See also Hamano (1986) for an alternative formulation in which accent is assigned on the syllable contained in the strongest foot within a prosodic word.

Many mimetic roots appear reduplicated, and in many such cases, the accent falls on the initial syllable, as in (49).

(49) Initially-accented reduplicated forms
  a. do'ki+doki 'nervous'
  b. mo'zi+mozi 'shy'
  c. go'ro+goro 'rolling'
  d. ba'ta+bata 'hectic'
  e. bu'ru+buru 'vibrating'
  f. ki'ra+kira 'shining'

Some other reduplicated forms are unaccented, as in (50).

(50) Unaccented reduplicated forms
  a. gaku+gaku 'quarrelsome'
  b. moku+moku 'quietly'
  c. tan+tan 'cooled down'
  d. yuu+yuu 'relaxed'
  e. men+men 'wide-spread'

In some instances, the same mimetic form can be initially-accented or unaccented, in which case (un)accentedness correlates with a particular semantic feature. When such forms are used adverbially to represent something ongoing, the forms tend to be accented; on the other hand, when the forms are used to represent a resultative state, the forms are unaccented (Tamori 1983), as some pairs in (51) show.

(51) Reduplicated mimetic forms with and without accent
  a. pi'ka+pika to hikaru 'flashes shiningly'
  b. pika+pika ni migaku 'to polish something shiny'
  c. tu'ru+turu to taberu 'eat smoothly (slurping)'
  d. turu+turu ni suru 'to polish something smooth'
  e. bo'ko+boko to sita miti 'a bumpy road'
  f. boko+boko ni suru 'to hit somebody and cause injury'

For further data and analysis involving the phonological and accentual properties of mimetics, see Hamano (1986) and Nasu (2002).

8 Interaction with other phonological phenomena

Accent interacts with many phonological processes in Japanese. This section provides a brief overview of how Japanese accent placement interacts with other phonological phenomena.

8.1 Epenthesis

Cross-linguistically, it is common to avoid placing stress – or metrical prominence in general – on epenthetic vowels (Alderete 1999a; Broselow 1982; Gouskova and Hall 2009). Evidence for this sort of avoidance is also found in Japanese. Kubozono (2001b, 2006, 2011) shows that in loanwords consisting of a light syllable followed by a heavy syllable (LH), accent falls on the initial syllable if the first vowel is not epenthetic, as in (52). Placing accent on the initial syllable is avoided, however, if the initial vowel is epenthetic, as in (53) (epenthetic vowels are shown by < >) (Kubozono 2011: 2887).

(52) Initial accent in LH if the first vowel is not epenthetic
 a. se'-dan 'sedan'
 b. ha'-wai 'Hawaii'
 c. de'-byuu 'debut'
 d. ka'-nuu 'kanoe'
 e. gi'-taa 'guitar'
 f. pu'-rin 'pudding'

(53) Final accent in LH if the first vowel is epenthetic
 a. t<u>-i'n 'twin'
 b. t<o>-ra'i 'try'
 c. d<o>-ra'i 'dry'
 d. g<u>-re'e 'grey'
 e. b<u>-ru'u 'blue'
 f. d<o>-ro'o 'draw'

As an interesting complication, it is not the case that accent on epenthetic vowels is simply prohibited altogether (e.g., /k<u>'rasu/ 'class' and /d<o>'resu/ 'dress'). It is only when two constraints are violated – (i) placing accent on epenthetic vowels, and (ii) placing accent on a light syllable in the presence of a following heavy syllable – that Japanese allows final accent in LH sequences. In this sense, this pattern constitutes a case of "a gang effect" where a phonological process happens only when two independently motivated phonological pressures are at work (Crowhurst 2011; Pater 2009; Smolensky 1995).

See also (17) above and Kubozono (2001b) for other potential cases of the accent-epenthesis interaction.

8.2 Rendaku

Another phonological pattern that interacts with accent is rendaku, voicing of initial consonants in the second members of compounds (see Vance, this volume). For

some (or most) morphemes, rendaku is optional, and Sugito (1965) points out that rendaku is often accompanied by deaccenting in family names, especially in those names that end with /+ta/. This contrast is illustrated by the examples in (54) and (55).

(54) Rendaku → unaccented
 a. yosi+da
 b. yama+da
 c. ike+da
 d. mae+da
 e. oka+da
 f. matu+da

(55) No-rendaku → accented
 a. hu’zi+ta
 b. mo’ri+ta
 c. si’ba+ta
 d. ku’bo+ta
 e. yo’ko+ta
 f. a’ki+ta

There are, however, some exceptions; e.g., /oo+ta/ and /ha’ra+da/ (Sugito 1965; Zamma 2005). Sugito (1965) and Zamma (2005) present quantitative surveys of names that end with this morpheme, which show the correlation between the presence of rendaku and unaccentedness, as in Table 6.

Table 6: Correlation between rendaku and accent. Reproduced from Zamma (2005: 159)

	accented	unaccented	either (variation)	total
no rendaku	94	13	10	117
rendaku	64	95	56	215
either (variation)	8	0	22	30
Total	166	108	88	362

This connection between rendaku and unaccentedness seems to be observed in other domains, including island names with /+sima/ (Tanaka 2005), last names ending with /+kawa/ (Ohta 2013), and the light verb /+suru/ (Kurisu 2010; Okumura 1984) (see also Yamaguchi 2011 and references cited therein); some examples are shown in (56).

(56) The interaction between Rendaku and accent
 a. awazi’+sima vs. sakura+zima ‘(place name)’
 b. okino’+sima vs. iriomote+zima ‘(place name)’
 c. yosi’+kawa vs. sina+gawa ‘(personal name)’

d.	nami'+kawa	vs.	uda+gawa	'(personal name)'
e.	tai+su'ru 'to oppose'	vs.	mee+zuru	'to order'
f.	koo+su'ru 'to resist'	vs.	hoo+zuru	'to report'

8.3 Vowel devoicing

Finally, accentuation interacts with high vowel devoicing in Japanese. Vowels between two voiceless consonants or those that are word-final and preceded by a voiceless consonant devoice in Japanese (see Fujimoto, this volume, for full discussion of vowel devoicing). There is a tendency to avoid placing accent on devoiced vowels, which is natural given that the metrical prominence provided by accent may not be very audible in voiceless vowels, since they do not involve robust periodic energy.

To illustrate this avoidance of accenting devoiced vowels, recall for example that accented verbs usually have their accent on the penultimate syllable – but, when the vowel in that syllable is devoiced, the accent can shift (Hirayama 1960; Vance 1987). This interaction between devoicing and accent shift is illustrated in (57). The example in (57d) (from Akinaga 1985: 8) shows this shift with an alternation: when the stem vowel gets devoiced, because the suffix-initial consonant is /t/, the accent may shift to the suffixal vowel.

(57) Accent shift due to devoicing
- a. tu̥'k+u → tu̥k+u' 'arrive'
- b. hu̥'k+u → hu̥k+u' 'blow'
- c. kaku̥'s+u̥ → ka'ku̥s+u̥ 'hide'
- d. hu'r+u 'fall' → hu̥t+te' 'falling'

In short, there is a tendency to shift accent due to vowel devoicing. However, young speakers place accent on devoiced vowels and show no such accent shifts. See Akinaga (1985), Kitahara (1996) and Maekawa (1990) and references cited therein for more on the interaction between vowel devoicing and accent.

9 Theoretical contributions

The discussion so far has been more or less descriptive, although the discussion also included basic metrical analyses of the Japanese accent system. Now we briefly turn to the theoretical contributions that Japanese accentology has made in the history of generative phonology.

Although it is not possible – or useful, even – to fully reproduce theoretical analyses of Japanese accent in various theoretical frameworks, it is probably important

to note that Japanese accentology has contributed to developments in phonological theory. Japanese accent has been analyzed from several theoretical perspectives throughout the history of generative phonology, starting from McCawley (1968). Readers are referred to the original references for the details of each analysis and the implications that it had for contemporary theoretical debates.

In early years of generative studies, attention was paid to the issue of how to represent Japanese accent phonologically. For example, Haraguchi (1977, 1991) developed autosegmental analyses (Goldsmith 1976) of accent patterns in many dialects in Japan, deriving surface tonal patterns from underlying diacritics using (universal) autosegmental conventions. Some other authors developed more purely tonal analyses without resorting to underlying lexical diacritics (Pierrehumbert and Beckman 1988; Poser 1984; Pulleyblank 1984). Pierrehumbert and Beckman (1988) moreover showed, based on experimental work, that "spreading of tones" (section 1.4) can be better analyzed as phonetic interpolation, building on the idea of phonetic underspecification (Keating 1988).

Within the framework of Metrical Phonology and Prosodic Phonology (Liberman and Prince 1977; Nespor and Vogel 1986; Selkirk 1978, 1980), in which linguistic utterances are organized into a set of hierarchical levels, Poser (1990) made an important contribution by showing that languages that do not possess stress (like Japanese) show evidence for the presence of a foot in their metrical organization. This contribution was not trivial because the foot was first proposed to compute stress placement (Hammond 2011; Hayes 1995; Selkirk 1980), and therefore it was not clear whether non-stress languages like Japanese could possess metrical feet or not.

Haraguchi (1999) offers an extensive analysis of the accentual behavior of verbs and adjectives (section 5) using the notion of extrametricality (Hayes 1982; Hyde 2011) and tonal spreading (Goldsmith 1976). The difference between the accentuation pattern of the non-past tense (penultimate) and the past tense (antepenultimate) has been analyzed in several ways, including extrametricality (Haraguchi 1999), the level ordering hypothesis (Clark 1986) (see Kiparsky 1982; Siegel 1974) and paradigm uniformity (Yamaguchi 2010) (see Benua 1997; McCarthy 2005a; Steriade 2000).

Soon after Optimality Theory (Prince and Smolensky 1993/2004) was proposed and became a dominant analytical framework in the field of phonology, Kubozono (1995, 1997) argued that a constraint-based model accounts well for various aspects of Japanese accent patterns. In particular, Kubozono (1995, 1997) developed a unified constraint-based analysis of compound accent rules, which we discussed in section 4.

Within Optimality Theory, the basic antepenultimate accent rule can be derived by having a trochaic foot with the final syllable unparsed (e.g., /kuri(su'ma)su/), and this foot placement can be explained as an interaction of two independently motivated constraints: RIGHTMOSTNESS and NONFINALITY, both of which have been proposed by Prince and Smolensky (1993/2004). The former constraint requires feet to be aligned to the right edge of a prosodic word, and one formulation of the latter

constraint requires that final syllables be unparsed. If NONFINALITY dominates RIGHTMOSTNESS, the final syllable remains unparsed, but the foot is placed rightwards to the extent possible.

Various morphologically controlled accent patterns have been analyzed from the perspective of modeling the phonology-morphology interface, especially in terms of how suffixes can affect roots' phonological shapes (Alderete 1999b, 2001a; Kawahara and Wolf 2010; Kurisu 2001; Inkelas 2011; Inkelas and Zoll 2007). For example, given that languages generally preserve underlying information from roots more often than from affixes, the behavior of dominant suffixes remains mysterious. Several solutions have been proposed to address this question; e.g., anti-faithfulness constraints (Alderete 1999b, 2001a) and a morpheme realization constraint (Kurisu 2001).

The privileged status of nouns – as compared to adjectives and verbs – in allowing contrastive accent locations has been discussed from the perspective of category-specific phonological patterns (Smith 1998, 2011). Cross-linguistically, there seems to be a general tendency to allow more contrasts in nouns than in adjectives and verbs, and Japanese fits this generalization well. Smith (1998) develops an analysis of this privileged status of Japanese nouns using category-specific faithfulness constraints.

10 Remaining issues

Accent is arguably the most extensively studied area in Japanese phonology, and the previous studies reviewed above, both in the traditional literature and in theoretical linguistics, have revealed many interesting patterns. There are a number of issues that remain to be addressed, however.

10.1 Experimentation with nonce words

Most studies on Japanese accent are based on descriptions in a dictionary (e.g., NHK 1998) or on impressionistic observations about existing words, and this chapter itself is no exception. This tradition is perhaps not without a reason – even linguistically naive native speakers of Tokyo Japanese have a fairly clear idea of accentual differences that exist among different words, and when asked, it is not difficult for them to choose an appropriate accent pattern for a particular word, even if they cannot identity its exact tonal contour or accent placement. Therefore, the data on Japanese accent, even though based on impressionistic observations, are fairly reliable.

Nevertheless, there have been a number of experimental works using nonce words (e.g., Katayama 1998; Kawahara and Kao 2012; Kubozono and Ogawa 2005; Tanaka 1995). Given the rise of laboratory approaches to phonology in recent years (Beckman and Kingston 1990; Pierrehumbert, Beckman, and Ladd 2000), we have

much to learn from experiments using nonce words (for example, wug-tests: Berko 1958). For example, experimentation is useful in order to address the true productivity of a particular accent pattern, or in order to solve particular theoretical debates, or to examine the quality of the data itself.

One could argue that loanword adaptation, reviewed in section 2, constitutes a more or less natural experiment on how Japanese speakers assign accent to nonce words (Kang 2011). However, when it comes to accentuation, one cannot deny the possibility that Japanese accent locations are influenced by the stress in the donor language (Akinaga 1985, cf. Kubozono 2006).[29] Studying loanword accentuation undoubtedly provides insight into Japanese accentuation systems, but systematic experimentation can complement that sort of study.

10.2 Lexical specification

As we have discussed throughout this chapter, Japanese does seem to have a default antepenultimate accent rule (or Latin Stress Rule). One question that arises is whether or not nouns that have lexical accent that happen to coincide with default accent (e.g., /i’noti/ ‘mind’) should be underlyingly specified for accent. A dominant assumption in the field has been that Japanese accent is unpredictable in nouns, so that accent locations need to be specified for all nouns. Kubozono (2006, 2008, 2011) challenges this traditional view, because if Japanese has a mechanism that assigns default accent, such lexical specifications are redundant. To the extent that Japanese phonology has a default accentuation assignment system, learners may as well take “a free ride” (McCarthy 2005b) on this rule, and leave the lexical representations unspecified.

This proposal is reminiscent of the idea that redundant features should be underspecified in the lexicon (i.e., the theory of underspecification) (e.g., Archangeli 1988). Although this theory has been challenged in Optimality Theory (McCarthy and Taub 1992; Prince and Smolensky 1993/2004), some recent psycholinguistic work argues that mental lexicons are indeed underspecified (Eulitz and Lahiri 2004; Lahiri and Marslen-Wilson 1991; Lahiri and Reetz 2002). On the other hand, there are some other lines of psycholinguistic work which argues for the opposite – that linguistic memories are richly encoded, including redundant information (e.g., exemplar theory) (Gahl and Yu 2006; Johnson 2007; Mitterer 2011).

This issue of underspecification is thus perhaps best discussed at two distinct levels – theoretical and psycholinguistic – and the Japanese accent system would bear on this debate from both perspectives. To address the question of whether

29 Some examples include /a’kusento/ ‘accent’, /fa’inansu/ ‘finance’, /ta’aminaru/ ‘terminal’, /sa’ikuringu/ ‘cycling’ and /sa’iensu/ ‘science’. These forms seem to reflect the stress pattern of the source language.

default accent is underspecified in the mental lexicon of actual Japanese speakers, psycholinguistic studies are necessary.

10.3 Acquisition of accent

As we have observed throughout this chapter, the Japanese accent system is a mixture of regularities and exceptions. To what extent semi-regular patterns are indeed grammaticalized in speakers' minds is an important but difficult question to address. One way to address this problem is to study the acquisition of accent patterns, and there have been various studies on this topic (Sato, Sogabe, and Mazuka 2010; Shirose, Kubozono, and Kiritani 1998; Shirose 2001). It is sometimes hard to transcribe accent patterns in child speech, but the study of acquisition of accent patterns (both in L1 and L2) should nevertheless provide us with much insight (see Ota, this volume, and Hirata, this volume).

11 Conclusion

This chapter has provided an overview of various aspects of Japanese accent patterns. It is impossible to provide a fully detailed description of the system in one chapter, let alone its analysis, so the aims of the current paper have been to introduce the basic patterns and analyses, and to place the discussion in cross-linguistic perspective. One of the challenges that Japanese accent patterns pose – which is an interesting one – is that the accent system of Japanese both show regularity and complex exceptions at the same time. The Japanese system will and should continue to provide an interesting testing ground for theoretical discussion.

Acknowledgements

I am grateful to Yosuke Igarashi, Shin-ichiroo Ishihara, Michinao Matsui, Toshio Matsuura, Yasutada Sudo, and Kazuko Yatsushiro for the contributions to the data collection. I am also grateful to Manami Hirayama, Yosuke Igarashi, Kazu Kurisu, Toshio Matsuura, Akio Nasu, Seunghun Lee, and Kyoko Yamaguchi for comments on earlier drafts of this chapter. Two reviewers (Haruo Kubozono and Tim Vance) provided very detailed comments on earlier drafts, which significantly improved this chapter. Donna Erickson, Jeremy Perkins, and Jess Trombetta very carefully proofread the drafts. All remaining errors are mine.

References

Akinaga, Kazue. 1985. Kyōtsūgo no akusento [Accent of standard Japanese]. In NHK (ed.), *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Akinaga, Kazue. 2002. Tokyo no hatsuon no yure [Fluctuation of Tokyo Japanese pronuciations]. In Yoshifumi Hida and Takeyoshi Sato (eds.), *Gendai nihongo kōza 3: Hatsuon* [Series in modern Japanese 3: Pronunciation], 40–58. Tokyo: Meiji Shoin.

Alderete, John. 1999a. Head dependence in stress-epenthesis interaction. In Ben Hermans and Marc van Oostendorp (eds.), *The derivational residue in Optimality Theory*, 29–50. Amsterdam: John Benjamins.

Alderete, John. 1999b. *Morphologically-governed accent in Optimality Theory*. Amherst MA: University of Massachusetts dissertation.

Alderete, John. 2001a. Dominance effects as transderivational anti-faithfulness. *Phonology* 18(2). 201–253.

Alderete, John. 2001b. Root-controlled accent in Cupeño. *Natural Language and Linguistic Theory* 19 (3). 455–502.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–208.

Becker, Michael and Kathryn Flack. 2011. The emergence of the unmarked. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1003–1026. Malden, MA and Oxford: Wiley-Blackwell.

Beckman, Jill. 1998. *Positional faithfulness*. Amherst, MA: University of Massachusetts dissertation.

Beckman, Mary. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, Mary and John Kingston. 1990. An introduction. In John Kingston and Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 1–16. Cambridge: Cambridge University Press.

Benua, Laura. 1997. *Transderivational identity: Phonological relations between words*. Amherst MA: University of Massachusetts dissertation.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14. 109–149.

Braver, Aaron. 2013. *Degrees of incompleteness in neutralization: Paradigm uniformity in a phonetics with weighted constraints*. New Brunswick, New Jersery: Rutgers University dissertation.

Broselow, Ellen. 1982. On predicting the interaction of stress and epenthesis. *Glossa* 16. 115–132.

Clark, Mary. 1986. Japanese as a tone language. In Takashi Imai and Mamoru Saito (eds.), *Issues in Japanese linguistics*, 53–105. Dordrechte: Foris.

Coetzee, Andries W. 2009. Learning lexical indexation. *Phonology* 26. 109–145.

Crowhurst, Megan. 2011. Constraint conjunction. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 3*, 1461–1490. Malden, MA and Oxford: Wiley-Blackwell.

Eulitz, Carsten and Aditi Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16 (4). 577–583.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Gahl, Susanne and Alan Yu (eds.). 2006. Special issue on exemplar-based models in linguistics. *The Linguistic Review* 23(3).

Giriko, Mikio. 2009. *Nihongo ni okeru go ninshiki to heiban-gata akusento* [Word structure recognition and the unaccented pattern in Japanese]. Kobe: Kobe University dissertation.

Giriko, Mikio, Takao Ohshita and Haruo Kubozono. 2011. Gomatsu ga zu dearu chīmumei/gurūpumei no akusento bunseki [On the nature of the morpheme [zu] in Japanese]. *NINJAL Research Papers* 2. 1–18.

Goldsmith, John. 1976. *Autosegmental phonology*. Cambridge, MA: MIT dissertation.

Gordon, Matthew. 2002. A phonetically driven account of syllable weight. *Language* 78(1). 51–80.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23. 595–653.

Gouskova, Maria and Nancy Hall. 2009. Acoustics of unstressable vowels in Lebeanese Arabic. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 203–225. London: Equinox.

Hamano, Shoko. 1986. *The sound-symbolic system of Japanese*. Gainesville, Florida: University of Florida dissertation [CSLI, 1998].

Hammond, Michael. 2011. The foot. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 949–979. Malden, MA and Oxford: Wiley-Blackwell.

Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.

Haraguchi, Shosuke. 1991. *A theory of stress and accent*. Dordrecht: Foris.

Haraguchi, Shosuke. 1999. Accent. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 1–30. Oxford: Blackwell.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13. 227–276.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: The University of Chicago Press.

Hirayama, Teruo (ed.). 1960. *Zenkoku akusento jiten* [All-Japan accent dictionary]. Tokyo: Tōkyōdō Shuppan.

Hoequist, Carles E. 1982. Duraional correlates of linguistic rhythm categories. *Phonetica* 40. 19–31.

Hulst, Harry van der. 2011. Pitch accent systems. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1003–1026. Malden, MA and Oxford: Wiley-Blackwell.

Hyde, Brett. 2007. Non-finality and weight-sensitivity. *Phonology* 24(2). 287–334.

Hyde, Brett. 2011. Extrametricality and non-finality. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1027–1051. Malden, MA and Oxford: Wiley-Blackwell.

Hyman, Larry M. 2009. How (not) to do phonologial typology: The case of pitch-accent. *Language Sciences* 31. 213–238.

Hirata, Yukari. this volume. Chapter 18: L2 phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Igarashi, Yosuke. this volume. Chapter 13: Intonation. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Inkelas, Sharon. 1999. Exceptional stress-attracting suffixes in Turkish: Representation vs. the grammar. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The prosody-morphology interface*, 134–187. Cambridge: Cambridge University Press.

Inkelas, Sharon. 2011. The phonology-morphology interaction. In John Goldsmith, Jason Riggle and Alan Yu (eds.), *The handbook of phonological theory*, 2nd edition. 68–102. Oxford: Blackwell.

Inkelas, Sharon, Orhan Orgun and Cheryl Zoll. 1997. The implications of lexical exceptions for the name of grammar. In Iggy Roca (ed.), *Derivations and constraints in phonology*, 393–418. Oxford: Oxford University Press.

Inkelas, Sharon and Cheryl Zoll. 2007. Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45. 133–172.

Inoue, Fumio. 1998. *Nihongo uocchingu* [Watching Japanese]. Tokyo: Iwanami.

Ishihara, Shin-ichiro. this volume. Chapter 14: Syntax-phonology interface. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Ito, Junko. 1990. Prosodic minimality in Japanese. In Michael Ziolkowski, Manual Noske and Karen Deaton (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.

Ito, Junko, Yoshihisa Kitagawa and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5. 217–294.

Ito, Junko and Armin Mester. 2003 [1992]. Weak layering and word binarity. In Takeru Homma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.), *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*, 26–65. Tokyo: Kaitakusha.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Oxford: Blackwell.

Ito, Junko and Armin Mester. 1996. Stem and word in Sino-Japanese. In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics*. Cambridge: MIT Press.

Ito, Junko and Armin Mester. 2007. Prosodic adjunction in Japanese compounds. *Proceedings of Formal Approaches Japanese Linguistics 4*, 97–112.

Ito, Junko and Armin Mester. 2008. Lexical classes in phonology. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 84–106. Oxford: Oxford University Press.

Ito, Junko and Armin Mester. 2012. Unaccentedness in Japanese. Ms. University of California, Santa Cruz.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Ito, Junko and Armin Mester. this volume. Chapter 9: Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Maria-Josep Solé, Patrice Beddor and Manjari Ohala (eds.), *Experimental approaches to phonology*, 25–40. Oxford: Oxford University Press.

Kang, Yoonjung. 2011. Loanword phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 1003–1026. Malden, MA and Oxford: Wiley-Blackwell.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology.* Sanra Cruz, California: University of California dissertation.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of sokuon, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. 2011. Experimental approaches in theoretical phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 2283–2303. Malden, MA and Oxford: Wiley-Blackwell.

Kawahara, Shigeto. this volume. Chapter 1: The phonetics of sokuon, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto and Sophia Kao. 2012. The productivity of a root-initial accenting suffix, [-zu]: Judgment studies. *Natural Language and Linguistic Theory* 30(3). 837–857.

Kawahara, Shigeto, Kohei Nishimura and Hajime Ono. 2002. Unveiling the unmarkedness of Sino-Japanese. In William McClure (ed.), *Japanese/Korean Linguistics 12*, 140–151. Stanford: CSLI.

Kawahara, Shigeto and Takahito Shinya. 2008. The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase and Utterance. *Phonetica* 65(1–2). 62–105.

Kawahara, Shigeto and Matthew Wolf. 2010. On the existence of root-initial-accenting suffixes: An elicitation study of Japanese [-zu]. *Linguistics* 48(4). 837–864.

Kawakami, Shin. 1961. On the relationship between word-tone and phrase-tone in Japanese Language. *The Study of Sounds* (The Phonetic Society of Japan) 9. 169–177.

Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 5. 275–292.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In In-Seok Yang (ed.), *Linguistics in the morning calm, vol. 2*, 3–91. Seoul: Hanshin.

Kisseberth, Charles. 1970. The treatment of exceptions. *Papers in Linguistics* 2. 44–58.

Kitahara, Mafuyu. 1996. Nihongo ni okeru museika onsetsu to picchi akusento no jitsugen ni tsuite [Devoiced syllables and the realization of pitch accent in Japanese]. *Gengo Kenkyu* 15. 1–17.

Kitahara, Mafuyu. 2001. *Category structure and function of pitch accent in Tokyo Japanese*. Bloomington, Indiana: Indiana University dissertation.

Kobayashi, Megumi. 2003. Tōkyōgo ni okeru keiyōshi akusento no henka to sono yōin [An examination of adjective accent changes in Tokyo Japanese]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 7(2). 101–113.

Kubozono, Haruo. 1995. Constraint interaction in Japanese phonology: Evidence from compound accent. In Rachel Walker, Ove Lorentz and Haruo Kubozono (eds.), *Phonology at Santa Cruz* (PASC) *4*, 21–38. Santa Cruz: Linguistics Research Center, UC Santa Cruz.

Kubozono, Haruo. 1996. Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin* (The Phonetic Society of Japan) 211. 71–82.

Kubozono, Haruo. 1997. Lexical markedness and variation: A nonderivational account of Japanese compound accent. In Brian Agbayani and Sze-Wing Tang (eds.), *Proceedings of West Coast Conference on Formal Linguistics*, 273–287. CSLI.

Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo. 2001a. Comments on "Privative tone in Bantu". In Shigeki Kaji (ed.), *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Japanese Accentology, and Other Topics'*, 259–265. Tokyo: ILCAA.

Kubozono, Haruo. 2001b. Epenthetic vowels and accent in Japanese: Facts and paradoxes. In Jeroen van de Weijer and Tetsuo Nishihara (eds.), *Issues in Japanese phonology and morphology*, 113–142. Berlin and New York: Mouton de Gruyter.

Kubozono, Haruo. 2006. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116. 1140–1170.

Kubozono, Haruo. 2008. Japanese accent. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 165–191. Oxford: Oxford University Press.

Kubozono, Haruo. 2010. Accentuation of alphabetic acronyms in varieties of Japanese. *Lingua* 120. 2323–2335.

Kubozono, Haruo. 2011. Japanese pitch accent. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 5*, 2879–2907. Malden, MA and Oxford: Wiley-Blackwell.

Kubozono, Haruo. 2013. Japanese word accent. *Oxford Bibliographies Online*.

Kubozono, Haruo. this volume. Introduction: Introduction to Japanese phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kubozono, Haruo and Yayoi Fujiura. 2004. Morpheme-dependent nature of compound accent in Japanese: An analysis of "short" compounds. *Phonological Studies* 7. 9–16.
Kubozono, Haruo, Junko Ito and Armin Mester. 1997. On'inkōzō kara mita go to ku no kyōkai: Fukugō meishi akusento no bunseki [The phonological boundary between the word and the phrase: An analysis of Japanese compound noun accentuation]. In Spoken Language Working Group (ed.), *Bunpō to onsei* [Speech and grammar], 147–166. Tokyo: Kurosio.
Kubozono, Haruo and Armin Mester. 1995. Foot and accent: New evidence from Japanese compound accentuation. Ms. University of California, Santa Cruz.
Kubozono, Haruo and Shinji Ogawa. 2005. Sutoraiki wa naze suto ka? Tanshuku to tango bunsetsu no mekanizumu [Why *sutoraiki* becomes *suto*? Mechanisms for truncation and word segmentation]. In Tsuyoshi Oishi, Tetsuo Nishihara and Yoji Toyoshima (eds.), *Gendai keitairon no chōryū* [Trends in modern morphology], 155–174. Tokyo: Kurosio.
Kubozono, Haruo and Satoshi Ohta. 1998. *On'in kōzō to akusento* [Phonological structure and accent]. Tokyo: Kenkyusha.
Kurisu, Kazutaka. 2001. *The phonology of morpheme realization*. Santa Cruz, California: University of California dissertation.
Kurisu, Kazutaka. 2005. Gradient prosody in Japanese. *Journal of East Asian Linguistics* 14. 175–226.
Kurisu, Kazutaka. 2010. Japanese light verb voicing as a connective morpheme. *Proceedings of WAFL 6*, 213–227.
Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.
Lahiri, Aditi and William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38. 245–294.
Lahiri, Aditi and Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 637–675. Berlin: Mouton de Gruyter.
Liberman, Mark and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.
Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti. 2002. X-JToBI: An extended J-ToBI for spontaneous speech. *Proceedings of the 7th International Congress on Spoken Language Processing*, 1545–1548.
Maekawa, Kikuo. 1990. Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese. *Proceedings of ICSLP 1990*, 517–520.
Martin, Samuel E. 1967. On the accent of Japanese adjectives. *Language* 43. 246–277.
Matsumori, Akiko, Tetsuo Nitta, Noriko Kibe and Yukihiko Nakai (eds.). 2012. *Nihongo akusento nyūmon* [Introduction to Japanese accent]. Tokyo: Sanseido.
McCarthy, John J. 2005a. Optimal paradigms. In Laura Downing, T. Alan Hall and Renate Raffelsiefen (eds.), *Paradigms in phonological theory*, 170–210. Oxford: Oxford University Press.
McCarthy, John J. 2005b. Taking a free ride in morphophonemic learning. *Catalan Journal of Linguistics* 4. 19–55.
McCarthy, John J. and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology*. 79–153.
McCarthy, John J. and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In Merce Gonzalez (ed.), *Proceedings of the North East Linguistic Society 24*, 333–379. Amherst, MA: GLSA.

McCarthy, John J. and Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *University of Massachusetts occasional papers in linguistics* 18, 249–384. Amherst: GLSA.
McCarthy, John J. and Alison Taub. 1992. Review of Carole Paradis and Jean-François Prunet eds., 1991. The special status of coronals: Internal and external evidence. *Phonology* 9. 363–370.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
McCawley, James D. 1977. Accent in Japanese. In Larry Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), 261–302. Los Angeles: USC.
Mester, Armin. 1990. Patterns of truncation. *Linguistic Inquiry* 21. 475–485.
Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.
Mitterer, Holger. 2011. The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance* 37(2). 496–513.
Myers, Scott. 1998. Surface underspecification of tone in Chichewa. *Phonology* 15. 367–391.
Nasu, Akio. 2002. *Nihongo onomatope no gokeisei to inritsu kōzō* [Word formation and metrical structure in Japanese onomatopoetic words]. Tsukuba, Japan: University of Tsukuba dissertation.
Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology.* Dordrecht: Foris.
NHK. 1998. *Nihongo hatsuon akusento jiten* [The Japanese language pronunciation and accent dictionary]. Tokyo: Nihon Hōsō Shuppan Kyōkai.
Nishiyama, Kunio. 2010. Penultimate accent in Japanese predicates and the verb-noun distinction. *Lingua* 120(10). 2353–2366.
Ohta, Satoshi. 2013. On the relationship between rendaku and accent: Evidence from -kawa/-gawa alternation in Japanese surnames. In Van de Weijer and Tetsuo Nishihara (eds.), *Current issues in Japanese phonology: Segmental variation in Japanese*, 63–87. Tokyo: Kaitakusha.
Okumura, Mitsuo. 1984. Rendaku. *Nihongogaku* 35. 89–98.
Ota, Mitsuhiko. this volume. Chapter 17: L1 phonology: phonological development. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Otake, Takashi. this volume. Chapter 12: Mora and mora timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Pater, Joe. 2000. Nonuniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17. 237–274.
Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035.
Pater, Joe. 2010. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 123–154. London: Equinox.
Pierrehumbert, Janet B. and Mary Beckman. 1988. *Japanese tone structure.* Cambridge: MIT Press.
Pierrehumbert, Janet B., Mary Beckman and Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Phillip Carr and Gerard Docherty (eds.) *Phonological knowledge: Conceptual and empirical Issues*, 273–303. Oxford: Oxford University Press.
Poser, William. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.
Poser, William. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105.
Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14(1). 19–100.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In Michael Ziolkowski, Manuela Noske and Karen Deaton (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 355–398.

Prince, Alan and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell. (Originally circulated in 1993 as ms. University of Colorado and Rutgers University).

Pulleyblank, Douglas. 1984. Autosegmental accent. *Proceedings of NELS*, 343–359.

Revithiadou, Anthi. 1999. *Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems*. Leiden: University of Leiden dissertation.

Revithiadou, Anthi. 2008. Colored turbid accents and containment: A case study from lexical stress. In Sylva Blaho, Patrik Bye and Martin Kraemer (eds.), *Freedom of analysis?*, 149–174. Berlin: Mouton de Gruyter.

Rosenthall, Sam and Harry van der Hulst. 1999. Weight-by-position by position. *Natural Language and Linguistics Theory* 17. 499–540.

Sato, Yutaka, Yuko Sogabe and Reiko Mazuka. 2010. Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience* 22(11). 2503–2513.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (ed.), *Nordic prosody, vol. 2*, 111–140. Trondheim: TAPIR.

Selkirk, Elisabeth. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11. 563–605.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shinohara, Shigeko. 2000. Default accentuation and foot structure in Japanese: Evidence from Japanese adaptations of French words. *Journal of East Asian Linguistics* 9. 55–96.

Shirose, Ayako. 2001. Nihongo ni okeru akusento no kakutoku [The acquisition of accentuation in Japanese]. *On'in Kenkyū* [Phonological Studies] 4. 97–102.

Shirose, Ayako, Haruo Kubozono and Shigeru Kiritani. 1998. The acquisition of compound accent in Japanese. In Robert H. Mannell and Jordi Robert-Ribes (eds.), *Proceedings of the 5th International Conference on Spoken Language Processing* (ICSLP '98), Sydney, Australia, 2671–2674. Australia: Australian Speech Science and Technology Association, Incorporated (ASSTA).

Sibata, Takesi. 1994. Gairaigo ni okeru akusento kaku no ichi [The location of accent in loanwords]. In Kiyoji Sato (ed.), *Kokugo ronkyū 4: Gendaigo hōgen no kenkyū* [Studies in Japanese 4: Studies in modern Japanese and dialects], 388–418. Tokyo: Meiji Shoin.

Sibata, Takesi and Ritei Shibata. 1990. Akusento wa dōongo o dono teido benbetsu shiuru ka [To what extent can word accent distinguish homophones?]. *Keiryō Kokugogaku* 17(7). 311–323.

Siegel, Dorothy. 1974. *Topics in English morphology*. Cambridge, MA: MIT dissertation.

Smith, Jennifer. 1998. Noun faithfulness: Evidence from accent in Japanese dialects. In Noriko Akatsuka, Hajime Hoji, Shoichi Iwasaki, Sung-Ock Sohn and Susan Strauss (eds.), *Japanese/Korean Linguistics 7*, 611–627. Stanford: CSLI.

Smith, Jennifer. 2007. Source similarity in loanword adaptation: Correspondence Theory and the posited source-language representation. In Steve Parker (ed.), *Phonological argumentation*. London: Equinox.

Smith, Jennifer. 2011. Category-specific effects. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 2439–2463. Malden, MA and Oxford: Wiley-Blackwell.

Smolensky, Paul. 1995. On the internal structure of the constraint component CON of UG. Talk presented at the University of California, Los Angeles.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In Michael B. Broe and Janet B. Pierrehumbert (eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

Sugito, Miyoko. 1965. Shibata-san to Imada-san: Tango no chōkakuteki benbetsu ni tsuite no ichi kōsatsu [Mr. Shibata and Mr. Imada: A remark on the perceptual discrimination of words]. *Gengo Seikatsu* 165. 64–72.

Sugiyama, Yukiko. 2012. The production and perception of Japanese pitch accent. Cambridge Scholars Publishing.

Tamori, Ikuhiro. 1983. Onomatope – On'in keitai to goisei – [Onomatopoeia–Phonological shapes and lexicality]. *Jinbun Ronshū* 19(2). 63–85.

Tanaka, Shin'ichi. 1995. Onsetsu kōzō kara mita katakana muimigo to gairaigo no akusento [Accent of nonce words and loanwords from the perspective of syllable structure]. *Nihongo/Nihon Bunka Kenkyū* 5. 59–68.

Tanaka, Shin-ichi. 2001. The emergence of the "unaccented": Possible patterns and variations in Japanese compound accentuation. In Jeroen van de Weijer and Tetsuo Nishihara (eds.), *Issues in Japanese phonology and morphology*, 159–192. Berlin: Mouton de Gruyter.

Tanaka, Shin-ichi. 2002. Three reasons for favoring constraint reranking over multiple faithfulness. *Shogengo no on'in kōzō to onsei riron ni kansuru sōgōteki kenkyū* [A comprehensive study of phonological structure and phonological theories] (KAKENHI report), 121–130.

Tanaka, Shin-ichi. 2005. Where voicing and accent meet: (In-)completeness in the harmonic scale of phonological prominence. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 261–278. Berlin and New York: Mouton de Gruyter.

Trubetzkoy, Nikolai S. 1939/1969. *Grundzüge der phonologie*. Güttingen: Vandenhoeck and Ruprecht (English translation by Christiane A.M. Baltaxe 1969. Principles of phonology. Berkeley: University of California Press).

Urbanczyk, Suzanne. 2006. Reduplicative form and the root-affix asymmetry. *Natural Language and Linguistic Theory* 24. 179–240.

Urbanczyk, Suzanne. 2011. Root-affix asymmetry. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 4*, 2490–2515. Malden, MA and Oxford: Wiley-Blackwell.

Uwano, Zendo. 1999. Classification of Japanese accent systems. In Shigeki Kaji (ed.), *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics'*, 151–186. Tokyo: ILCAA.

Uwano, Zendo. 2007. Two-pattern accent systems in three Japanese dialects. In Carlos Gussenhoven and Tomas Riad (eds.), *Tones and tunes, vol. 1: Studies in word and sentence prosody*, 147–165. Berlin: Mouton de Gruyter.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: SUNY Press.

Vance, Timothy J. 1995. Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics* 23(4). 487–499.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.

Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Venditti, Jennifer. 2005. The ToBI model of Japanese intonation. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 172–200. Oxford: Oxford University Press.

Warner, Natasha. 1997. Japanese final-accented and unaccented phrases. *Journal of Phonetics* 25. 43–60.

Yamaguchi, Kyoko. 2010. The difference in accentuation between the present and the past tenses of verbs in Japanese. *Journal of the Phonetic Society of Japan* 14(3). 1–10.

Yamaguchi, Kyoko. 2011. Accentedness and rendaku in Japanese deverbal compounds. *Gengo Kenkyu* 140. 117–134.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yu, Alan. 2011. Contrast reduction. In John Goldsmith, Jason Riggle and Alan Yu (eds.), *The handbook of phonological theory*, 2nd edition, 291–318. Oxford: Blackwell.

Zamma, Hideki. 2005. Correlation between accentuation and rendaku in Japanese surnames: A morphological account. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 157–176. Berlin and New York: Mouton de Gruyter.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12. 85–129.

Takashi Otake

12 Mora and mora-timing

1 Introduction

The present chapter aims to provide a comprehensive review of mora-timing based upon an understanding of the concept of mora. Mora and mora-timing are linguistic terms to denote separate notions. Since the notion of mora-timing depends on the definition of "mora", different accounts of mora-timing can emerge.

The mora is a phonological unit, with a rich history, used to describe lexical organization characterized by two kinds of syllables: short (or light) and long (or heavy). It has the specific role to describe the internal structure of syllables in terms of "weight" (see Broselow 1995 for the weight system and Blevins 1995 and Bosch 2011 for a review of the internal structure of syllables). The traditional weight system is dichotomously defined: a light syllable (CV syllable) is associated with one moraic unit, while a heavy syllable (CVV or CVC syllables) is associated with two moraic units. What is light or heavy, however, is not so simple according to present phonological theory (Zec 2011). Syllable weight is linked with prosodic phenomena such as stress assignment or rhythmic organization of naturally spoken languages, as we will see in section 3.

Mora-timing is a term which appeared in the mid-twentieth century in the literature to describe a particular type of speech rhythm (see Warner and Arai 2000 for a review of mora-timing). This speech rhythm is characterized such that the mora is the recurring unit (e.g., Bloch 1950; Hockett 1955; Ladefoged 1975), just as syllable and stress are recurring units in syllable- and stress-timed languages, respectively (e.g., Pike 1945). Mora-timing today is being challenged because several views of "a recurring unit" are proposed in the literature. These views may be summarized as follows (see Kubozono 1999 and section 5 below for more details): (i) mora as a timing unit (Han 1962; Homma 1981; Port et al. 1987), (ii) mora as a boundary unit (Otake et al. 1993; Cutler and Otake 1994; Cutler and Otake 2002), and (iii) mora as an invariable syllabic unit (Dauer 1983; Ramus et al. 1999; Low et al. 2000; Nespor et al. 2011).

The present chapter has two tasks. The first task is to present an in-depth review of the notion of mora. We examine how the notion of mora emerged, what properties were associated with it, and how different views were assigned to it. Since our central focus is limited to mora-timing, we will not go into the detail of the notion of mora used in theoretical phonology (see Broselow 1995 and Blevins 1995). The second task is to review the notion of mora-timing. We examine how different views of mora are associated with mora-timing.

This chapter is organized as follows. Section 2 sketches issues on mora-timing from a historical perspective, including stress- and syllable-timing. Section 3 reviews the notion of mora including how the mora emerged in the analysis of classical and modern languages. Section 4 reviews issues about the mora in the context of Japanese, including how and when the term was brought into the discussion of Japanese. This is followed by section 5, where we review three interpretations of mora-timing. Section 6 discusses other issues related to mora-timing, followed by a concluding section where future issues on mora-timing are raised.

2 Overview: speech rhythm and mora-timing

The emergence of a global world has a significant impact on the investigation of speech rhythm today because researchers can have easy internet access to almost all spoken languages on the earth. This newly available technology is particularly beneficial for two reasons for those who investigate speech rhythm. First, a wide variety of spoken languages can be accessible instantly with great ease, so that auditory impressions of speech can be easily judged. Second, and more importantly, researchers can discharge a volley of questions to native speakers directly about whether their judgments are correct or not. This follow-up test is vital for deciding speech rhythm because it cannot be determined without examining its cognitive aspect. As Pike (1947: 65) clearly claimed, “Phonemic analysis cannot be made with the phonetic data alone; it must be made with phonetic data plus a series of phonemic premises and procedures.” His observation is important for the examination of speech rhythm, too. It can be proven only when the researchers’ judgment based upon speech signal is correctly validated by native speakers’ intuitions. However, this procedure was not necessarily appreciated fully by the researchers, as will be shown below.

2.1 Speech rhythm

When a speaker expresses his or her ideas using spoken language, various prosodic properties inevitably accompany the sequence of segments. Speech rhythm is one of the prosodic properties. Since the physiological gestures of the speech organs in the course of utterances vary from one speaker to another, theoretically speaking, identical acoustic cues for speech rhythm do not exist in human speech. In spite of this fact, different languages sound like they have different speech rhythms. Now one may wonder what constitutes speech rhythm.

Over half a century ago some structural linguists hypothesized that all spoken languages could be classified into two types of speech rhythm, stress-timing and syllable-timing (Pike 1945; Abercrombie 1967). Pike (1945: 35) heard a small number

of languages (English and Spanish) and proposed that the auditory impression of these languages differed such that stressed syllables recurred at roughly equivalent intervals in English, while syllables recurred at roughly the same interval in Spanish. He cited the common labels proposed by Lloyd (1940) to characterize the auditory impression of the utterances of these languages: "Morse code" for stress timing and "machine-gun" for syllable timing. "Morse code" refers here to a succession of both short and long sounds, while "machine-gun" refers to a rapid succession of short sounds. Shortly afterwards, mora-timing was proposed as a third category of speech rhythm on the basis of a single language, Japanese. Bloch (1950: 90) pointed out that "the auditory impression of any phrase is of a rapid patterning succession of more-or-less sharply defined fractions, all of about the same length." Bloch (1950: 91) argued that "all these fractions are heard as having the same time value." He proposed the label "staccato" to describe mora-timing.

2.2 Mora-timing and syllable-timing

One may think that this brief historical sketch casts serious doubt on mora-timing because similar labels ("machine gun" and "staccato") were used to describe the distinctively different categories of speech rhythm. These two labels denote almost the same meanings in the sense that an event recurs at the same time interval. One may wonder why almost the same labels were used for different categories of speech rhythm and what makes the distinction between them. Pike (1945: 35) argued that the recurrence of the prosodic unit in Spanish speech might be the syllable. Bloch (1950: 89), on the other hand, claimed that the recurrence of the prosodic unit in Japanese speech might be a mora.

The question here is how by Spanish or Japanese speakers can distinctively recognize syllable and mora in the speech stream to determine speech rhythm. The significant point that Bloch assumed was that the acoustic signal of Japanese speech itself consists of a succession of moras, each of which is about the same length and that this acoustic signal must be responsible for the auditory impression of a staccato rhythm (Bloch 1950: 90). To see whether this assumption is tenable, let us examine a simple case.

Suppose one is viewing Japanese automobile commercial films produced by car dealers (*Toyota*, *Subaru*, *Honda*, *Nissan, etc.*) both in Tokyo and Madrid. The viewer's task is to evaluate the auditory impression of speech rhythm for the brand names while watching the commercials on YouTube. Her auditory impression would probably be "machine gun" or "staccato" for both languages, as Pike and Bloch reported. Note that according to the speech rhythm hypothesis, the acoustic properties of these words uttered by Spanish and Japanese speakers must reflect both speech rhythms. Then, she should ask herself to determine which term, syllable or mora, would describe what she is actually hearing. If she is a Spanish speaker, or for that

matter, any non-native speaker of Japanese, she may perceive *Toyota* and *Subaru* as trisyllabic words (*To.yo.ta* and *Su.ba.ru*) and *Honda* and *Nissan* as bisyllabic words (*Hon.da* and *Nis.san*) regardless of whether it is a Spanish or Japanese language commercial. On the other hand, native speakers of Japanese will perceive the former two words as trimoraic words (*To.yo.ta* and *Su.ba.ru*) and the latter two words as trimoraic (*Ho.n.da*) and tetra-moraic (*Ni.s.sa.n*) words, respectively, regardless of speech materials. The question is why this different interpretation emerges among different language users. The answer is that some language users recognize syllables as phonemic syllables and other language users, as phonetic syllables, as pointed out by Pike (1947: 65). Thus, it may be that Japanese listeners interpret the speech stream as a series of moras, while Spanish listeners as a series of phonetic syllables. Both language users may perceive the speech stream in different ways, i.e., mora for Japanese speakers and syllable for Spanish speakers (Otake et al. 1993; Otake et al. 1996). This may imply that Japanese listeners recognize speech signals without referring to the durational property (i.e., a temporally equal duration unit) because the Spanish speech signals do not meet the acoustic information required for mora-timing. Thus, this may suggest that the durational property of mora may not be the only property attributable to mora-timing. One may wonder then what property of mora listeners perceive to hear mora-timing in speech. Up to now, three hypotheses have been proposed in the literature of phonetics and psycholinguistics.

The first hypothesis is that the mora is a timing unit which always has the same duration (Bloch 1950: 90), as described above. According to this hypothesis, spoken words in utterances consist of chunks (moras), each of which is equal in length. It should be noted that this assumption could be applied to long syllables as well, suggesting that long syllables may consist of two moraic units, as proposed in classical languages (see section 3.1), and are two times longer than short syllables (Bloch 1950; Hockett 1955; Ladefoged 1975).

The second hypothesis is that the mora is a boundary unit in speech segmentation (Otake et al. 1993; Cutler and Otake 1994; Cutler and Otake 2002). According to this hypothesis, any utterances can be divided into chunks (moras) on the basis of a mora boundary. The important difference between the first and second hypotheses is that the former assumes that both short and long syllables are recognized by the length of mora, while the latter assumes that they are recognized by the mora boundaries because the number of moras and mora boundaries in those syllables are identical.

The third hypothesis is that the mora is the least variable syllable unit (Dauer 1983; Ramus et al. 1990; Nespor et al. 2011). It is important to note that this hypothesis makes a distinction among syllable-, stress-, and mora-timing by the degree of variability of syllables (Ramus et al. 1999) and that the properties of the mora described above are irrelevant.

3 What is a mora?

Let us begin by asking what a mora is. As described in section 1, a mora is a unit that measures syllable weight. However, mora itself has been viewed in different ways by different researchers. In this section we review how mora was viewed by researchers during three different linguistic periods. The main concern here is to elucidate what properties of mora were proposed and which property was attributed to mora-timing.

The first period was the classical language period when mora emerged in the prescriptive grammar more than twenty centuries ago to describe the rhythmic organization of the verse form and to explain the location of accent in classical languages (see Donaldson 1848 and Allen 1973 for a full review). The second period was the structural linguistics period in the middle of the twentieth century when mora was used to describe prosodic properties of naturally spoken languages including non-European and Japanese languages (Bloch 1950; Hockett 1955; Trubetzkoy 1969). The third period was the generative linguistics period after the 1960s when mora was used to describe prosodic properties of modern languages with advanced theories of phonology (see detailed discussions in Hayes 1995, Hyman 1985, and McCarthy and Prince 1986).

3.1 Mora in classical languages

3.1.1 Emergence of mora

The structure of the lexicon in classical languages was composed of two types of phonemes: (i) quality and (ii) quantity (Allen 1973). Quality refers to the quality of speech segments, usually described by the manner and place of articulation, while quantity refers to the length of speech segments, mainly vocalic segments. As a result, the lexicon in the languages included qualitatively equivalent but quantitatively different words although not all languages contain both types. For example, two bisyllabic Latin words, *mălus* 'not good' and *mālus* 'apple tree' are such words where the two diacritics of the first syllables indicate a short and a long vowel, respectively. The former has a CV syllable, while the latter has a CVV syllable.

When the initial syllables of these words were classified in terms of the length of the vocalic element, the former (*mă-*) was classified as a short syllable, while the latter (*mā*) as a long syllable. This classification was based upon the length of the vocalic element. As for the second syllable (*-lus*) of *mālus*, the whole syllable must be called a short syllable since the vocalic element of this syllable is obviously a short vowel. However, an alternative new term was used to classify this type of syllable as a long syllable. This new term was called the "weight" system. In this

system, both a CVV syllable (*mā*) and a CVC (*lus*) were called a heavy syllable, while a CV syllable (*mă*) was a light syllable. In order to measure the weight of the syllable, another new term, mora, was used. By definition, a light syllable contained a one-mora weight, while a heavy syllable contained a two-mora weight. It should be noted that these things were determined "by convention" in classical languages.

Donaldson (1848: 16) stated that "the quantity of syllables is determined either by the nature of vowel, or by that of consonants which follow: in the former case the quantity is said to depend on the nature of vowel; in the latter, on the position of the consonants". This indicates that the physical vocalic length was not the only criterion to determine the weight of syllables (Allen 1973: 89). Once this extended interpretation was established, both CVV and CVC syllables were treated equally as heavy syllables in classical languages. Here a new question emerges as to why the classical languages devised the weight system in the first place. This is discussed in the following two sub-sections.

3.1.2 Mora and rhythmic organization

We now look at how mora was utilized in classical languages. First, it was used to organize the verse form in these languages. The verse system of these languages was characterized by the alternation of the two types of syllables, light and heavy. Halporn et al. (1963: 4) noted that "the rhythm of classical Greek poetry is determined by the 'flow' or succession of long and short elements." What this statement means is that whenever lexical items were arranged in verse form, they observed particular rhythmic patterns created by the two types of syllables: short and long. According to Donaldson (1848: 16), "the shortest time in which a syllable can be pronounced is called a mora, or a single time. A short syllable has one mora: a long syllable contains two morae". Thus, given the words *mălus* (short + long) and *mālus* (long + long) in the Latin verse, even though both were bisyllabic words, the former was read as a three-mora, while the latter read as a four-mora word.

The typical canonical patterns preferred in classical languages were short + long (the iambus), long + short (the trochee), long + short + short (the dactyl), etc. (Bingham, 1871: 328–329). Thus, the word *mălus*, which combines short and long syllables, could create the iambic rhythmic pattern, while the word *mālus* did not create any desirable patterns because the sequence of two consecutive long syllables was not regarded as canonical.

One significant aspect of mora was that, as Donaldson (1848) states, each mora had to be pronounced with equal duration because mora was regarded as a timing unit. Thus, when *mălus* (short + long) was pronounced, the second syllable took two times longer than the first one. Thus, it was often said that the ratio between a light and a heavy syllables had to be 1:2 (Bingham 1871: 318). Allen (1973: 46), however, remarked about light and heavy syllables in versification in a different way: "It has

sometimes been assumed that, as the terminology suggests, the distinction is simply one of temporal duration. But phonetic studies of spoken languages in which this distinction is made show that, whilst ‘long’ vowels do tend to be of longer duration than ‘short’, and normally are so in comparable environments, the actual durations fluctuate to a considerable degree, and it is doubtful whether the hearer could always use them as sole criteria for judging the category to which a particular vowel sound belongs. Moreover, the relationship of the perceptual dimension of ‘length’ to objective duration seems not to be a simple one, and is not yet fully understood.” If Allen’s comment is correct, physical duration is unlikely to be a primary property of mora in the verse form of classical languages. It should be added here that the statements in grammars of classical languages were regarded as prescriptive rather than descriptive.

In sum, the first function of mora in the classical verse was “a measuring tool” to organize the rhythmic patterns in poetry. The basic process of composing poetry was to select the best suited lexical items which met the canonical patterns from a list of possible candidate words.[1] This may indicate that the phonetic value of short and long is a secondary matter for poets.

3.1.3 Mora and accent assignment

Now let us consider the second function of mora in classical languages. Typologically speaking, classical languages are said to be quantity-sensitive languages, in that the location of accent is predictable by the nature of syllables (Zec 2011). The accent was placed on a syllable located just before the last one of a word and was sensitive to the “quantity” of the syllable, i.e., short (light) or long (heavy) syllable. For example, if the penultimate syllable was a long syllable, accent was placed on this syllable. If it was short, accent was placed on the preceding syllable. Trubetzkoy (1969: 174) explained the location of word accent using mora in the following way (see Kawahara Ch. 11, this volume, and Kubozono Ch. 8, this volume, for the resemblances between Latin and Japanese accent):

> The same interpretation is also given to long nuclei in those languages where length in the delimitation of words is treated according to the formula “one long unit = two short units”. Classical Latin may be cited as a generally known example, where the accent delimiting words could not fall on the word-final syllable. It always occurred on the penultimate “mora” before the last syllable, that is, either on the penultimate syllable, if the latter was long, or on the antepenultimate syllable, if the penultimate was short. A syllable with a final consonant was considered long. A long vowel was thus comparable to two short vowels or to a “short vowel + a consonant”.

1 Moseley (1827) published a dictionary of short and long syllables. He mentioned in the preface that ‘The author having examined every syllable in the Latin language, and found, that with few exception, they are both long or short, whether Final, Middle or Initial.’

In sum, mora was utilized as “a measuring tool” to determine the length of syllables because it was helpful to identify the location of accent in classical languages.

3.2 Mora in the structural linguistics period

The structural linguists both in Europe and North America revived the notion of mora through the investigation of the prosodic system of non-European languages spoken in Africa, North and South America and Asia toward the middle of the twentieth century (Trubetzkoy 1969; Trager 1941; Pike 1947; Hockett 1955). They found that the notion of mora was very helpful to describe the prosodic system of non-European languages (Trubetzkoy 1969: 182). In this section we look at how the notion of mora was viewed to describe the prosodic system of naturally spoken languages and to describe mora-timing.

3.2.1 Syllable nuclei and mora: imaginary and real languages

Trubetzkoy (1969) argued that many of the prosodic systems of non-European languages including Japanese assigned prosodic properties (stress, tone and pitch) on syllables rather than vowels, whereas he assumed that vowels were the basic unit for the prosodic properties of European languages (Trubetzkoy 1969: 171). Following this observation, he proposed to use the term “syllable nuclei” for the domain of prosodic properties. The significant point here is that any syllable constituent could be a candidate for bearing a prosodic unit. These constituents were (i) a vowel, (ii) a combination of vowels, (iii) a consonant, and (iv) a vowel and a consonant (Trubetzkoy 1969: 170). A number of structural linguists followed Trubetzkoy’s observation and gave much supportive evidence using examples from both imaginary and real languages.

Pike (1947) illustrated how syllable nuclei should be dealt with in terms of prosodic units by demonstrating a specific analytical procedure with an imaginary language called *Kataba*. Pike (1947: 145) demonstrated a syllable nucleus comprised of a long vowel in this language. He assumed that three types of distinctive words existed in this language: (i) [tótò] ‘tomato’, (ii) [tóà] ‘corn’ and (iii) [tâ] ‘potato’. (i) is a bisyllabic word involving a sequence of two CV syllables, where a high and a low tone are assigned on the first and second syllables, respectively. (ii) shows two consecutive vowels in syllable nuclei in which the two tones are assigned on the two separate vowels. (iii) illustrates a long vowel in which the two tones, a high and a falling tone, are assigned onto the vowel. Pike assumed that all these words consist of two moras because two tone values were assigned to them. His analysis in the imaginary language illustrates mora as an accent-bearing unit, which is an abstract unit that can be defined without reference to the durational property.

What about real languages? Trager (1941: 136) described mora as "an abstract unit which was assigned to short, long and diphthongs such that one mora for a short vowel and two moras for a long vowel or diphthong" (see Kubozono Ch. 5, this volume, for diphthongs in modern Japanese). This observation was based upon Navaho in which the two basic tones (high and low) were assigned on a long vowel or diphthong, i.e., the vocalic elements were treated as two elements, just as in *Kataba*. Thus, the analytical procedure was basically the same as the one used by Pike.

Hockett (1955: 61) referred to the durational property to explain the notion of mora. He mentioned that "A syllable containing two vocoids lasts about twice as long as a syllable containing one. We may therefore introduce the term mora: a syllable contains one or two moras." Here, unlike Trubetzkoy and Trager, Hockett made an association between the syllable length and the durational property of mora.

Hockett (1955: 61) further referred to the durational property of mora in his analysis of nasals in Chiricahua. He reports that 'there are also syllables with no onset or coda consisting of the consonant /n/ or /m/ with one of the two tones; or of the geminate cluster /nn/ or /mm/, each with one of the two tones.' This indicates that the geminate nasals in Chiricahua receive two separate tones like /ḿm̀/, so that the sequence of two nasals has to be regarded as two mora units. It is important to note here that he assumed that the two consecutive nasals in a syllable nuclei were twice as long as the single case.

Hockett (1955: 54–55) provides further evidence from different languages to support his argument. For example, he mentioned that "In Senadi, the essential difference between a two-peak sequence /áà/, the first with low tone and the second with high, and a single peak /ǎ/ with two successive tones, lies in the fact that the former takes approximately the same length of time as a sequence like /tàká/, while the latter takes approximately the same length of time as /à/ or /á/." Notice that the logic behind this explanation is the same as the one by Pike (1947) described above. He analyzes sequences of nasals in Bariba and long vowels in Fijian using exactly the same logic.

3.2.2 Syllable nuclei and mora: Japanese

Hockett (1955) accounts for the two types of syllable nuclei (CVV and CVC) in Tokyo Japanese in terms of mora using the same procedure described just above. Given the three bisyllabic words, *koo.ka* 'a coin', *kon.do* 'next time,' and *kok.ki* 'a national flag,' the prosodic property (pitch) is assigned to the syllable nuclei as in HL.L, HL.L and LH.H, respectively (see Kawahara Ch. 11, this volume, for details about Japanese word accent). In these words, the two pitch values, H and L, are placed on the initial

syllable nuclei (HL for *koo*- and *kon*- and LH for *kok*-), so that these syllables consist of two moras, according to the analytical procedure.

Now consider another case, *nippon* 'Japan,' which is a bisyllabic word containing two syllable nuclei. The accent pattern for this word is LH.HL, suggesting that two pitch values are assigned to each syllable nuclei. The two pitch values are placed on both syllable nuclei (LH for *nip*- and HL for -*pon*), so that each syllable nuclei consists of two moras. In referring to the durational property of mora for this Japanese word, Hockett (1955: 59) states :

> /nippon‸/ 'Japan' takes just about the same length of time to utter as does /sayonara/ 'goodbye' where all the syllables contain onset and peak; the syllables of the first are /ni/ (onset plus peak), /p/ (acoustically a silence of approximately syllabic duration), /po/ (onset plus peak) and /n‸/ (syllabic nasal). We cannot class this Japanese system in any of the other three types (peak, onset-peak, or onset), because the Japanese syllable is defined fundamentally in terms of duration and nothing else.

The logic behind this assumption seems to be the same as the one mentioned in section 3.2.1.

Bloch (1950) claimed for the first time that Japanese speech rhythm is mora-timing. Approaching the subject from a different perspective, he proposed that the Japanese lexicon is organized by a unit of duration, or mora. Bloch (1950: 90–91) stated:

> the number of syllables in a phrase is therefore not found by counting peaks of sonority or chest pulses, but only by counting the temporally equal fractions contained in it, or by comparing its duration with that of another phrase in which the number of fractions is known. In short, the Japanese syllable is a unit of duration. Such a unit is called a mora.

Bloch apparently claimed that Japanese perceive the unit of duration by counting its number or by comparing its duration. It is interesting to note that Bloch (1950: 92) also mentions the possibility that Japanese speakers count the boundary between syllables although he did not propose how to count the boundary within long syllables. This alternative idea seems to be more realistic for determining the mora because Japanese syllables are typically CV syllables.

In sum, Trubetzkoy (1969) and Trager (1941) referred to the structural property of mora, but not to the durational property, while Pike (1947) and Hockett (1955) referred to both the structural and durational properties. On the other hand, Bloch (1950) referred only to the durational property of mora.

3.3 Mora in the generative linguistics period

One of the significant contributions on mora by the generative linguists was that they advanced theoretical phonology with the notion of mora to analyze wide variety

of languages. However, our concern here is not to explore their contributions, but to review mora-timing. Specifically, we will review the notion of mora proposed by McCawley (1968, 1978) because the two views he expressed in his works seem to reflect how the generative linguists perceived the notion of mora.

McCawley (1968) argues that the notion of mora has to be used to describe Tokyo Japanese for three reasons. First, the mora is to be used as a unit of length, as proposed by his predecessors (e.g., Bloch 1950; Hockett 1955). That is, the length of phrases or words is roughly proportional to the number of moras they contain (see the argument in Bloch 1950: 92). Moreover, Japanese meters are characterized by the template of 5 or 7 moras, suggesting that the principle above is applied to the meter system. Second, Japanese accent is described by mora as an accent-bearing unit. For example, *kokusai* 'international' and *kooban* 'police box' consist of the tonal string LHHH. Notice that the initial syllable of the latter word is a long vowel pronounced as LH, which indicates that two tonal values are assigned on it. Third, some phonological rules such as the loanword accent rule are based upon moras (see Kubozono Ch. 8, this volume, for details about loanwords and their accent patterns). Mora as described in these ways is a durational-based or an accent-bearing unit and does not refer to the distinction between short (light) or long (heavy) syllables as discussed in the previous sections. This suggests that McCawley's (1968) claim is very similar to the one proposed by the structural linguists. However, McCawley later suggests as follows (McCawley 1978: 114):

> There is only one workable universal definition of "mora": something of which a long syllable consists of two and a short syllable of one. That is, a long syllable can be divided into something of the shape of a short syllable plus something else, and both of these are moras.

In the later generative linguistic period the notion of mora came to be considered an important prosodic unit. Several notable phonologists such as Hayes (1995) and Hyman (1985) argue that mora plays a significant role in the Moraic Theory (see Broselow 1995 for a general review). The significant point about the notion of quantity is basically abstract rather than physical (Odden 2011: 465), suggesting that the notion of mora can be attributed to a structural property without reference to its durational property.

3.4 Summary

In this section we looked at how the notion of mora was viewed during three periods of linguistic research. There were several notable features on the notion of mora with respect to durational and structural properties. First, in the classical language period, the distinction between short and long syllables was originally made by the physical vocalic length. In order to treat the two types of long syllables (CVV and

CVC) under the same category, the weight system was introduced and mora was proposed to measure the weight of syllables. Although mora was associated with the abstract unit, it was also used as a unit to measure the physical length or duration in pronunciation. This suggests that both structural and durational properties were associated with mora. Second, both properties were taken over by some structural linguists such as Pike and Hockett but not by others such as Trubetzkoy and Trager. Another type of mora which was solely associated with the durational property was proposed, too (e.g., Bloch). Third, both properties were taken over by some generative linguists, but the durational property was abandoned (McCawley 1978). Since then, mora has been associated with only the structural property.

4 Two kinds of prosodic units: mora and *haku*

In section 3 we looked at the notion of mora from the Greco-Roman period to the generative linguistics period, focusing on the properties that were attributed to mora. However, traditional Japanese linguists utilized a somewhat alternative term of mora called *haku* 'beat' whose literal meaning denotes that it recurs at regular intervals. In this section, we will look at several topics related to this traditional unit.

4.1 Japanese lexicon and syllable structure

One of the characteristics of Japanese lexicon from a historical perspective is that there were dual stages regarding syllable structure in the lexicon. Prior to the 11th century, all Japanese words were constructed by light syllables (i.e., V and CV) only. For example, *biwa* 'a Japanese lute', *kokoro* 'mind', *murasaki* 'purple' were all native words in old Japanese, each of which was made up of only light syllables. After this period, however, new words with heavy syllables (i.e., CVV and CVC) emerged in Japanese for two historical reasons, one was phonological changes called *onbin* and the other, lexical borrowings from Chinese words (Komatsu 1977; Okimori 2010: 140). For example, *yonde* (a gerund form of *yomu* 'to read') and *motte* (a gerund form of *motu* 'to have') emerged because of the phonological change. *Konro* 'a fireplace' emerged as a new word borrowed from Chinese. Furthermore, it is said that CCV syllables (the second C being limited to the glide /j/) also emerged because of the language contact with Chinese (Okumura 1972 : 87).

It is important to note that although the two types of syllables exist in Japanese today, the distribution of these syllables in the lexicon is heavily skewed. For example, Otake (1990) reported that 70% of syllables in spontaneous conversation in Japanese are light syllables. Kubozono (1995) reported similar results based upon a dictionary search. These facts suggest that even though two types of syllables exist in the

Japanese lexicon, heavy syllables in this language may not have be dealt with in the same way as in classical languages.

Given the dual stages of syllables in the history of Japanese and the heavily skewed distribution of syllables in modern Japanese, it is interesting to know how Japanese syllables have been treated in the *haku* system and how mora was introduced in modern phonology. We look at these two problems below.

4.1.1 Haku and mora in Japanese

The term mora in the modern phonology of Japanese was introduced by several linguists such as Hattori (1960). However, *haku* was also used by Japanese linguists such as Kindaichi (1967). These two terms were proposed to make the distinction between phonemic and phonetic syllables (Kamei 1956).

In modern Japanese, there are two types of *haku*. The first type is called a *jiritsu haku* 'independent *haku*' which is equivalent to light syllables (i.e., V, CV or CCV). For example, *biwa* 'a Japanese lute', *kokoro* 'mind' and *murasaki* 'purple' are made up of two, three and four *jiritsu haku*, respectively. Most of the single *jiritsu haku* can be used as independent words, i.e., monomorphemic words. For example, *me* 'an eye', *i* 'a stomach' and *te* 'a hand' are all monomorphemic words. The second type of *haku* is called a *tokushu haku* 'special *haku*', which is equivalent to the second element of heavy syllables (i.e., -V in CVV, -Q in CVQ and -N in CVN, where Q and N denote moraic obstruents and nasals, respectively).

There is a notable feature specifically characteristic of *tokushu haku*. Heavy syllables were always treated as two separate units, a *jiritu haku* plus a *tokushu haku*, rather than a single unit as represented in the classical languages. As a result, the second element of heavy syllables (-V, -Q and -N in CVV, CVQ and CVN) had an explicit status, indicating that there were two *haku* units in heavy syllables. Thus, native speakers of Japanese share the knowledge of this explicit status within heavy syllables. For example, *honda* is not represented as *hon.da*, but as *ho.n.da* in the *haku* system, suggesting that there are two syllable boundaries in the former and three *haku* boundaries in the latter including the boundary coinciding with the word-final boundary. In this respect, the internal structure of heavy syllables in Japanese is explicitly represented in the *haku* system, while that of heavy syllables in classical languages was implicitly represented in the syllabic system.

In sum, although the notion of mora may be compatible with that of *haku* in the sense that there are two units in heavy syllables, the latter is treated as an explicit unit, while the former is not.

4.1.2 Haku and the writing system

In the preceding section, we looked at how heavy syllables in Japanese are represented in the *haku* system. As is widely known, the Japanese writing system is a

syllabary system which is compatible with the *haku* system. In this section, we look at the main characteristics of the Japanese writing system (see the chapters in the History Volume for full discussion).

A writing system did not exist in Japanese until the 7th century. There were two stages for developing a basic writing system in Japanese. At the first stage prior to the 11th century, Chinese characters whose sounds were similar to Japanese light syllables were used to represent native Japanese words. For example, the native Japanese word *yama* 'a mountain' was represented as 夜麻 because the sounds of these Chinese characters 夜 and 麻 were *ya* and *ma*, respectively. (Recall that prior to the 11th century, all Japanese words were composed of light syllables. Today, *yama* is represented by a single (bisyllabic) Chinese character 山.) This strategy was extremely productive, but it was confusing because there were too many homophonous characters in Chinese. Furthermore, a Chinese character consisted of many strokes, so that it took too much time to write a single Chinese character to represent each light syllable. Thus, it was desirable to develop a simpler writing system.

At the second stage a new simplified set of syllabary symbols were developed from Chinese characters whose sounds were similar to those in Japanese, either by deforming the character or extracting part of it. For example, the Chinese character 以, whose sound was *i*, was deformed to い. Thus, the syllabary い corresponds to /i/ in Japanese. *Hiragana* letters were formed in this way. An example of "extracting" is イ coming from the Chinese character 伊, which was pronounced as /i/ and underwent the deletion of the right-hand element. This type is called *katakana*. These two kinds of kana syllabary systems were developed to represent Japanese words for light syllables.

When heavy syllables emerged after the 11th century, how were they represented in the Japanese writing system? As we have seen, heavy syllables were interpreted as a combination of a *jiritsu haku* plus *tokushu haku*. When *tokushu haku* emerged, new kana syllabary symbols were also created. These are a small tsu, っ (ッ) for the moraic obstruent and ん (ン) for the moraic nasal, where the symbols in the parentheses are *katakana* letters. As a result, heavy syllables came to be represented by two symbols in Japanese. For example, *Honda* and *Nissan* can be represented as ホンダ (three kana syllabaries) and ニッサン (four kana syllabaries). Notice that the number of *haku* is basically the same number of *kana* syllabary items, implying that recognition of *haku* may be deeply associated with the orthographic representation of Japanese.

4.2 Haku and classical Japanese poetry

According to *The Oxford English Dictionary*, poetry is defined as a "literary work in which the expression of feelings and ideas is given intensity by the use of distinct style and rhythm." This definition implies that poetry in different languages and

cultures may develop a wide variety of strategies to express its own distinct style and rhythm reflecting, among others, that the structure of the lexicon varies from one language to another.

In section 3.2 we saw that the rhythmic structure of the classical verse system was manifested by the alternation of both light and heavy syllables. We also saw that the mora was exploited as an index to measure the rhythmic patterns within or between words. In this section we look at how the rhythmic structure of classical Japanese poetry was represented by the *haku* (or mora) system.

The rhythmic structure of classical Japanese poetry (called *tanka*) was realized by syllabic meter. Although this meter system was widely exploited in a variety of languages such as French and Chinese, the Japanese meter system was different from the systems of those languages in that the lexicon exploited in classical Japanese poetry was limited to native Japanese words with only light syllables (i.e., V or CV).

As a result, each line of classical Japanese poetry was represented by a sequence of *haku*: there were five lines, each of which was organized by either five or seven *haku* units. It is obvious that this kind of syllable arrangement cannot create the rhythmic structure preferred in classical languages. Chamberlain (1887: 4–5), a 19th century scholar with a great store of knowledge about poetry both in European and Asian languages, described the characteristics of classical Japanese poetry in the following way (underline by the present author).

> Of all such complication Japanese prosody knows nothing. It regards neither rhymes, tone, accent, <u>quantity</u>, nor alliteration, nor does its rather frequent parallelism follow any regular method. <u>Its only essential rule is that every poem must consist of alternate lines of five and seven syllables to mark the close. It is, indeed, prosody reduced to its simplest expression.</u> Yet so little artifice is needed to raise prose to verse in this most musical of tongues, that such a primitive metre still satisfies the native ear to-day in every street-ballad, as it already did in the seventh century Mikado's court; no serious attempt has ever been made to alter it in the slightest degree, even during the period of the greatest intellectual ascendancy of China.

There are a couple of important remarks in this citation. First, as the underlined part in the second line indicates, classical Japanese was not considered to be a quantity language. This is because the lexicon of classical Japanese was constructed by light syllables only (Komatsu 1977; Vance 1987: 56). Second, as the second underlined part in the third line shows, classical Japanese poetry used a fixed syllable-based template – five or seven *haku* (light syllables) – in each line as a "meter". The poem in (1), which was contributed by Abe-no Nakamaro to *Kokinshū* during the 10th century, has five lines, each of which is composed of either five or seven *haku* (light syllables, either V or CV).

(1) あまのはら (a-ma-no-ha-ra) (5)
ふりさけみれば (hu-ri-sa-ke-mi-re-ba) (7)
かすがなる (ka-su-ga-na-ru) (5)
みかさのやまも (mi-ka-sa-no-ya-ma-mo) (7)
いでしつきかも (i-de-shi-tsu-ki-ka-mo) (7)
(translation: I gaze across the / endless plains of the sky can / that moon be the one/that comes from the rim of Mount / Mikasa in Kasuga)[2]

Unlike the verse system in classical languages, the lexical items in classical Japanese poetry were always represented by *kana* syllabary as shown above. (It is said that Japanese classical poetry making was essentially orthography-based.) Or the composition of classical Japanese poetry was accomplished by allocating the lexical items which met the fixed number of *haku* template (five or seven light syllables) in each line. This suggests that there was no idea of measuring the weight of syllables in classical Japanese poetry (Kawamoto 1990: 233). It is widely believed by researchers that each *haku* is read in such a way that the length of each syllable is identical (Bekku 1977: 48; Kawamoto 1990: 221).

So far, we have looked at the structure of *tanka* poetry which is composed of light syllables. Basically the same is true of *haiku* poetry, which emerged during the 17th century. *Haiku* has a new poetry system whose template is shortened to 5-7-5 from the 5-7-5-7-7 template. This new type of Japanese poetry made it possible to use lexical items with heavy syllables. The *haiku* in (2) was composed by Matsuo Basho during the 17th century: Heavy syllables are underlined.

(2) てんびんや (te-n-bi-n-ya) (5)
きょうえどかけて (kyo-o-e-do-ka-ke-te) (7)
ちよのはる (chi-yo-no-ha-ru) (5)
(translation: On the scales/ Kyoto and Edo balanced/ in this spring of a thousand years)[3]

The first line contains two heavy syllables (CVC), while the second line contains one heavy syllable (CVV). Notice that there are two *kana* representations for each heavy syllable in (2), which indicates that the *haku*-based (or *kana*-based) representation played a key role in Japanese poetry. Moreover, even though two types of syllables (short/light or long/heavy) exist in Japanese, it never results in producing an alternation of these syllables. Besides, the number of heavy syllables in Japanese lexicon is very limited (Bekku 1977: 25; Kawamoto 1990: 233). It is important to note here that

2 Adapted from *Kokinwakashū* database (http://etymology.jp/gomit-the-db/KW/html/KW000406.html created by Hirofumi Yamanoto.
3 Adapted from Barnhill David Landis (2004) *Basho's Haiku: selected poems by Matsuo Basho*, State University of New York Press: Albany.

a heavy syllable appeared to be twice as heavy as a light syllable. However, this system is different from that of classical languages because there are two visible independent units in Japanese. Thus, the main strategy for poetry-making in classical Japanese was to select the best suited lexical item from a list of possible candidate words that satisfied both the meaning and the fixed template in order to create its own rhythmic effect.

4.3 Summary

In this section we looked at how the Japanese lexicon was represented by two kinds of prosodic units (*haku* and mora) and by the orthographic system. Three points must be noted here. First, the Japanese lexicon was traditionally represented by *haku* and mora. Second, *haku* and mora were not interchangeable prosodic units because the heavy syllables were analyzed as two separate units in the *haku* system but not in the moraic system. Third, native speakers of Japanese were fully aware of the explicitness of the system because the orthographic system was almost perfectly compatible with the *haku* system. We also looked at Japanese classical poetry whose metric system was governed by *haku*, suggesting that the Japanese verse system was different from the classical verse system in involving no weight system.

5 What is mora timing?

The general theory of speech rhythm has been primarily concerned with the question of what constitutes a recurring unit in the speech stream (section 2.1). The earliest researchers such as Pike (1945, 1947) and Abercrombie (1967) proposed that the recurring unit is a timing unit, e.g., the duration of syllables for syllable-timing and the duration between stressed syllables for stress-timing. Mora-timing has been added as a third type of speech rhythm under the same assumption (e.g., Bloch 1950). Thus, researchers who pursued mora-timing mainly investigated the physical duration of mora as a timing unit (Han 1962; Homma 1981; Port et al. 1987). As we saw in section 4, however, McCawley (1978) abandoned this assumption and claimed instead that syllable structure determines the status of mora in such a way that heavy syllables have two moras and light ones have one mora. In sum, the earliest researchers focused on the durational property of mora, while the later researchers paid attention to its structural property. Accordingly, mora-timing can be interpreted in several ways. In this section we review three hypotheses about this notion proposed in the literature.

5.1 Mora-timing as a timing unit

The first type of mora-timing hypothesis refers to the mora as a constant timing unit, as proposed by Bloch (1950). It assumes that the duration of mora plays a central role as a recurring unit in speech rhythm.

5.1.1 Rationale

The fundamental assumption of this hypothesis is that the mora recurs as a timing unit in speech stream, as we saw in section 3. There is one premise in this hypothesis. That is, since each mora has an equal duration, speech rhythm of mora-timing is solely determined by the number of moras contained in each lexical item. If a lexical item is composed of a sequence of light syllables, the speech rhythm becomes a staccato rhythm because each light syllable has an identical duration. Thus, Japanese automobile companies, *Toyota* and *Subaru* have the same rhythmic pattern because each word consists of three light syllables. Furthermore, *Honda* shows the same speech rhythmic pattern because it also consists of three moras, i.e., one light (one mora) and one heavy syllable (two moras). In other words, mora-timing is defined solely by the number of moras in lexical items.

5.1.2 How data were collected and interpreted

Although the debate on speech rhythm itself is crucially related to speech perception (recall the original hypothesis in section 2), researchers attempted to measure the acoustic signals read by Japanese subjects. We look at how data were collected in Han (1962), Homma (1981) and Port et al. (1987).

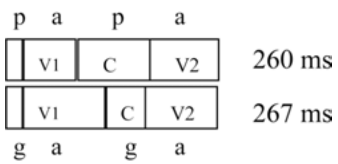
The experimental design to test the mora-timing hypothesis was rather simple at the earliest stage. The experimental materials were represented in *kana* or Roman letters with or without a carrier sentence. For example, Han (1962) used a sequence of light syllables such as /kakikukekokane/ (Japanese syllabary line with extra moras) or pairs of words like /obasan/ 'aunt' vs. /obaasan/ 'old woman' and /supai/ 'spy' vs. /suppai/ 'sour'. Homma (1981) used pseudo words like *kaka/gaga* with a carrier sentence. Port et al. (1987) used both real and non-words with a sequence of light syllables or heavy syllables embedded in a carrier sentence. The experimental material was designed such that a mora was added to the base sequence (for example, *ra* → *raku* → *rakuda* → *rakudaga* → *rakudagashi*). Furthermore, an additional set of devoiced materials was included in Port et al.'s experiment.

Once the materials were recorded by Japanese subjects, the relative duration of segments (consonants and vowels) and the duration of consonants and vowels in short and heavy syllables were measured.

Han (1962) found that vocalic and consonantal durations varied significantly between segments such that high vowels (/i/ and /u/) are always shorter than low

vowels (/a/ and /o/), voiceless fricatives are always longer than stops, voiceless stops are longer than voiced ones, etc. A similar tendency is also reported in English (e.g., Klatt 1976). While the segmental durations of vowels and consonants vary from one segment to another, the combined durations of vocalic and consonantal segments tend to be constant: for example, the vowel /a/ is shorter when it is combined with a voiceless stop, e.g., /p/, than when combined with other consonants. Han also found that although the durations of vocalic and consonantal segments in heavy syllables (for example, /obasan/ ‘aunt’ vs. /obaasan/ ‘old woman’ for vowels and /supai/ ‘spy’ vs. /suppai/ ‘sour’ for consonants) were not twice as long as those in light syllables, the durations of the moraic portion of a long vowel and a geminate consonant were equivalent to the preceding light syllables (see Kawagoe, this volume, and Kawahara Ch. 1, this volume, for full discussion of geminate consonants in Japanese). Han interpreted these results as evidence in support of mora-timing. She suggested that a timing control mechanism may be responsible for this phenomenon.

Following Han (1962) and Port et al. (1980), Homma (1981) argued for a temporal compensation effect that regulates the durations of adjacent vowels and consonants in order to keep moraic duration constant. She hypothesized that given bimoraic CV sequence like *papa/gaga*, a compensation effect would occur within the word-medial -VC- sequence as shown in (3), where /a/ is shorter when it is followed by /p/ than when followed by /g/.

(3)


As can be seen in (3), the word duration for these words is almost the same. However, the duration of the initial mora is considerably longer for *gaga* than *papa*. She showed that a compensation effect was observed between V_1 and C, namely, that temporal compensation worked not within a mora or a syllable, but within a word.

Following the findings by Han (1962) and Homma (1981), Port et al. (1987) investigated the durational property of mora-timing in a more elaborate way. Conducting several experiments, they discovered that the accumulated duration of words was directly proportional to the number of moras involved, as predicted in section 5.1. To sum, they claimed that there is a timing unit or a mechanism to regulate moraic duration.

5.2 The mora-timing as a mora boundary unit

A second type of mora-timing hypothesis is based on the notion of mora as a boundary unit and posits that speech rhythm in Japanese can be accounted for by counting the

number of mora boundaries as a recurring unit. It assumes that the moraic boundary plays a central role as a recurring unit in this type of speech rhythm. This hypothesis is based upon the notion of mora proposed in phonology.

5.2.1 Rationale

The fundamental assumption of this hypothesis is that mora recurs as a boundary unit in speech stream. According to the definition of mora proposed by McCawley (1978), light and heavy syllables contain one and two mora boundaries, respectively. For example, *Toyota* consists of three light syllables, so that there are three boundaries (To.yo.ta.). Honda also has three boundaries (Ho.n.da.) since it consists of one heavy and one light syllables. To put it another way, as long as one can recognize the number of mora boundaries, she can count the number of recurring units in the speech stream. Phonological knowledge may be associated with mora-timing, while a durational property is irrelevant, as McCawley (1978) pointed out. In order to support this hypothesis, it is vital to show evidence that heavy syllables are recognized as two moraic units in natural speech.

5.2.2 How data were collected and interpreted

The data collection for this hypothesis was derived from studies initiated at the beginning of the 1980s about word boundaries in spoken language. Mehler et al. (1981) hypothesized that syllables may play a central role as cues to word boundaries because all words in spoken languages can be represented by syllables. They tested this hypothesis with French listeners using French materials. They used an experimental task called a syllable monitoring task. French listeners heard a train of spoken French words and were asked to detect the beginning of a word with a target syllable, using a reaction time device. For example, the subjects were presented with spoken words beginning with two types of syllables, a light syllable /ba-/ (balance) and a heavy syllable /bal/ (balcony) and were asked to judge which word matched with the given target /ba/ or /bal/. The results showed that the judgment was faster and more accurate when the syllable type of the target and the initial syllable of the spoken word matched. Mehler et al. (1981) interpreted these results as evidence that French listeners use syllables to segment continuous speech.

Otake et al. (1993) hypothesized that Japanese listeners may segment continuous speech in Japanese with moras. In order to test this hypothesis, they used Japanese materials that are as similar as possible to the French materials. Recall that the French experiment used French materials *balance* and *balcon*, each of which contained *bal*- and *ba*- word-initially and targets */bal/* and */ba/* to test the matching process. The Japanese words *tansi* 'a terminal' and *tanisi* 'a snail' were chosen with targets /tan/ and /ta/. When the target /ta/ was presented, Japanese subjects

matched it with the word-initial CV mora of both *tansi* ‘a terminal’ and *tanisi* with the same accuracy and reaction time. When the target /tan/ was presented, only *tansi* was matched. Otake et al. interpreted these responses as evidence for mora-based segmentation.

It is important to note that the segmentation strategy between the two language groups differed from each other. For example, Japanese listeners segmented CV in CVCV- (*tanisi*) and CV in CVCCV- (*tansi*) as CV.CV and CV.CCV-, respectively. Furthermore, the initial CV of CVCV- was not segmented as CVC, but the initial CVC of CVCCV was segmented as CV.C. The results show that the Japanese listeners are aware of mora boundaries within heavy syllables, while French listeners are not aware of the internal boundary within a heavy syllable. Thus, if syllable-timing is closely associated with syllable boundaries, mora-timing is directly associated with mora boundaries in Japanese.

5.3 Mora-timing as the least variable unit

A third version of the mora-timing hypothesis assumes that speech rhythm can be accounted for by phonological factors. It assumes that mora-timing and syllable-timing differ from each other with respect to the degree of durational variability of vocalic and consonantal segments. There are two notable studies: Dauer (1983) and Ramus et al. (1999).

5.3.1 Rationale

Dauer (1983) argues that the dichotomous nature of speech rhythm (syllable-timing and stress-timing) itself can be explained if several phonological factors are taken into account (see also Dasher and Bolinger 1982, who proposed this argumentation for the first time). The main factors are syllable structure and vowel reduction. For example, higher preponderance of open syllables with little vowel reduction may cause an auditory impression of syllable-timing, while higher preponderance of closed syllables with a high degree of vocalic variation may cause an auditory impression of stress-timing. Dauer (1983: 60) proposes that speech rhythm of different languages may be defined on a continuous uni-dimensional scale, allocating mora-timing at the most right edge (the least variation), stress-timing at the most left edge (the maximum variation). She suggests that syllable-timing may be allocated between them. Although mora-timing was originally proposed in reference to the properties of mora, this factor is totally abandoned in her proposal, implying that speech rhythm is solely determined by the phonetic factors governed by the phonological elements.

Building upon Dauer's claim, Ramus et al. (1999) propose an insightful hypothesis. They argue that if speech rhythm is determined by phonetic factors, the acoustic signals may reflect different degrees of variability in the three kinds of speech rhythm and that infants may utilize the acoustic information in acquiring their native languages. In order to verify this hypothesis, they measured the durations of vocalic and consonantal intervals in languages which were categorized in one of the three speech rhythms. Specifically, they proposed three kinds of indices (Ramus et al. 1999: 272): (i) the proportion of vocalic intervals within the sentence which was denoted as %V, (ii) the standard deviation of the duration of vocalic intervals within each sentence which was denoted as ΔV, and (iii) the standard deviation of the duration of consonantal intervals within each sentence denoted as ΔC. They assumed that these indices are responsible for the different rhythm types (see Low et al. 2000, too, who proposed a similar system using different indices).

5.3.2 How data were collected and interpreted

Dauer (1983) conducted an experiment reading a passage from a novel or play by five different language speakers (English, Thai, Spanish, Greek and Italian), some of which were identified as stress-timing or syllable-timing while others were unidentified. The main goal of this experiment was to test whether or not the inter-stress intervals were manifested by different language speakers in accordance with the typology of speech rhythm. The results showed no difference in inter-stress intervals between English (stress-timing) and Spanish speakers (syllable-timing). Furthermore, speakers of other languages showed no more regularity than English speakers, either. She concluded that the inter-stress intervals were not reliable enough to determine speech rhythm.

Dauer assumed instead that variations in syllable structure might be closely related with the typology of speech rhythm. She pointed out as follows (Dauer 1983: 55):

> Over half of the syllables in Spanish and French have a simple CV structure, whereas in English there is a wider distribution among different kinds of syllables. Open syllables clearly predominate in both Spanish and French. We may have an impression of more regularity in a language such as Spanish or French from the frequent repetition of structurally similar open syllables.

She also remarked, "In addition to the greater variety of syllable structures typically found in a stressed-timed language, there is also a strong tendency for "heavy" syllables to be stressed and "light" syllables to be unstressed." Dauer examined these two factors and found out that her hypothesis can be largely supported. According to Dauer (1983: 56), the percentage of open syllables in all occurring syllables in Spanish, French and English was 70%, 74% and 44%, respectively. Heavy syllables in English tend to be stressed more than in Spanish: see Dauer (1983: 57) for more details.

In contrast, Ramus et al. (1999) measured both vocalic and consonantal segments in multi-lingual recordings in English, Polish, Dutch, French, Spanish, Italian, Catalan and Japanese. Speech rhythm of these languages can be classified into the three categories (stress-, syllable- and mora-timing). Their experiment was designed to test whether the acoustic measurements were directly correlated with the three categories of speech rhythm. The results supported their prediction: see Ramus et al. (1999: 273) for more detail. The most interesting result for our purpose is that the measurements of Japanese showed a very distinctive characteristic. That is, %V in Japanese was more than 50%, while it was around 40% for English, Dutch and Polish, and around 45% for Spanish, French, Italian and Catalan. This implies that the alternation of CV syllables occurred frequently in Japanese. They concluded that one may be able to predict the rhythm type of a language by looking at its values for ΔC and %V.

6 Issues on mora-timing

As we saw in section 5, the central question on mora-timing is which property of mora is associated with mora-timing. We looked at the three factors which are associated with mora-timing hypotheses: (i) mora as a timing unit, (ii) mora as a boundary unit and (iii) mora as an invariable durational unit. All these hypotheses have merits and demerits. We discuss some remaining issues related to these hypotheses here.

6.1 Mora as a timing unit

The central feature of the first hypothesis is that the the duration of moras must be constant in a mora-timed language. In order to maintain this feature, the duration of each mora must be regulated by the speaker. This hypothesis is subject to several controversies.

First, this hypothesis has been tested not by how moraic units in utterances are perceived by the speakers of a mora-timed language, but by how those units are produced by those speakers. This is probably because there was an assumption that a perceptual unit is equivalent to a production unit. Recall, however, that the general hypothesis on speech rhythm was originally to quest why utterances in Japanese are perceived as a staccato rhythm. Many researchers have attempted to demonstrate how uniquely a moraic duration is produced or regulated so as to be felt as a constant durational unit by the native speakers or in comparison with the acoustic signals in stress-timed or syllable-timed languages (see Warner and Arai 2000 for the full review).

Second, since moraic duration can be affected by various factors, two versions of mora-timing hypothesis have been proposed: a strong version and a weak version. The compensation effect discussed by Homma (1981) and Port et al. (1987) supports the weak version. This has helped to explain why the length of mora is regulated to

maintain a constant duration. But it is not clear how exactly it contributes to the auditory impression of mora-timing. Furthermore, this compensation mechanism was proposed mainly by the analysis of the acoustic signals uttered by Japanese speakers, not based on data from other languages which are also considered to be mora-timed languages. If another language which is claimed to be a mora-timed language shows the same effect, it may be strong evidence. Tamil is said to have mora-timing (Balasubramanian 1980, 1981; Nespor et al. 2011), but no substantial data have been reported.

Third, Bloch (1950) first argued that word duration was proportional to the number of moras within the word in a mora-timed language. This proposal was the original source of evidence for mora-timing. However, we must also remember that Bloch was one of the structural linguists. As we have seen in section 3, the structural linguists' main concern about prosodic properties was which unit, syllable or mora, served as an accent-bearing unit (e.g., Trubetzkoy 1969). Pike (1947) and Hockett (1955) proposed that once mora was treated as an accent-bearing unit, the durational property of mora had to be one unit or a half unit of a syllable. However, they never explained why it was so. One possible reasoning is that they might have dealt with the durational property in the same way as the prescriptive grammarians did at the classical language period.

In fact, there is another way to explain Bloch's argumentation described above. In section 4 we discussed the traditional *haku* system in Japanese, which strongly reflects the historical development of syllables in Japanese. The most notable feature of this system is that all heavy syllables are treated as if there were two independent units (i.e., *jiritsu haku* and *tokushu haku*). Since an independent status was given to *tokushu haku* not only in the writing system, but also in the phonological system for language games, each *haku* was treated as an independent single unit. Furthermore, since the basic syllable structure of *haku* in Japanese is very simple and equivalent to light syllables (e.g., V, CV and CCV), the duration for each *haku* is homogeneously similar (Beckman 1982). This simplicity of the *haku* system enables us to claim that word duration is proportional to the number of *haku*. Thus, in fact what Bloch argued might have referred to the *haku* system rather than mora.

Finally, let us consider the durational property of heavy syllables in the periods of classical languages and structural linguistics. Recall that heavy syllables were two times as long as light syllables in the classical languages because they had two moras. Furthermore, the same reasoning was also applied to the relationship between light and heavy syllables in Pike's (1947) and Hockett's (1955) treatments. The question is whether the utterances which contained heavy syllables and adjacent light syllables show a partial mora-timing in the classical languages or in languages where the accent is assigned based on mora. The answer will probably be a negative one since heavy syllables in those languages simply had the function of attracting the accent. Thus, the arguments based upon a constant timing unit seem to be misleading for mora-timing.

6.2 Mora as a boundary unit

One of the main arguments for the hypothesis of the mora as a boundary unit is that mora boundaries determine mora-timing. Thus, it does not assume that recognition of mora requires the durational property described above.

The main argument in the discussion of the weight system in section 3 was that there was one mora in light syllables and two moras in heavy syllables. There was no argument regarding mora boundaries. The weight system was used for specific purposes such as the organization of the rhythm, the location of the accent and classification of languages with respect to the accent assignment both in the classical language and the structural linguistics periods.

Now let us suppose that native speakers of mora-counting language have been asked to determine the mora boundaries within the car brand names such as *Toyota*, *Honda* and *Nissan*. We anticipate no one except Japanese speakers would show us clear answers because whether one can recognize mora may be task-oriented in nature. For example, speakers of classical languages such as Classical Greek and Latin may count the number of moras within words as long as their task is to organize rhythm in verse or to determine the location of the accent. Native speakers of Fijian or Bariba may also correctly assign the accent on moras in heavy syllables, but they may not be able to tell us the number of mora boundaries.

On the other hand, native speakers of Japanese – from four or five years old and above – can correctly tell the number of moras and the location of mora boundaries in the given car brand names. In fact, the number of moras and that of mora boundaries are identical. Japanese speakers can tell the number of moras and the location of mora boundaries because they have the knowledge of mora boundaries. They are aware of mora boundaries and know how to use them in the segmentation of continuous speech. (Otake et al. 1993; Cutler and Otake 1994; Otake et al. 1996; Inagaki et al. 2000; Cutler and Otake 2002). All native speakers of Japanese are able to read Japanese poetry (both *tanka* and *haiku*) because they know how to count the number of moras and mora boundaries. Japanese is rich in language games based on mora boundaries (Katada 1990; Haraguchi 1996).

Furthermore, the Japanese writing system (kana syllabary) mostly coincides with moras and mora boundaries. For example, when Japanese children start learning kana syllabary at elementary school, they use a specific notebook which contains a number of grids. Whenever a child writes a word in kana syllabary, she is instructed to write each kana symbol (both *jiritsu haku* and *tokushu haku*) in each grid box. The only exception is CCV moras, which are represented by two kana letters: Children are instructed to use two grid boxes in this exceptional case. In this way, children learn how to represent a word with letters as well as the number of moras and mora boundaries. In this sense, mora boundaries are explicit information.

If mora-timing is determined by the explicit knowledge on mora boundaries, the next question is which languages can be mora-timed. Marty et al. (2007) tested

Telugu, which was traditionally classified as a syllable-timed language, but showed a moraic effect in continuous speech (This effect was observed only in a native material (Telugu), not in nonnative material (Japanese)). Moreover, this language uses a syllabary-based writing system which is similar to that of Japanese. This may suggest that as long as explicit knowledge on mora boundaries is acquired in the phonological system, it is likely to be classified as mora-timed language. More languages need to be tested in future studies.

6.3 Mora as an invariable syllabic unit

A third hypothesis of mora-timing is based on the notion of mora as an invariant syllabic unit. It challenges the previous two hypotheses in that it does not presuppose the knowledge of mora for mora-timing. Dauer (1983) put forward doubts about the earliest premise regarding the dichotomy of speech rhythm (syllable-timing and stress-timing). She argued that different languages exhibit different auditory impressions largely because of differences in phonological structure such as the complexity of syllable structure and the presence of vowel reduction. As an alternative idea, she proposed a unidirectional scale where the three types of speech rhythm could be allocated on a language depending upon the phonological factors. For example, stress-timed languages are placed at the leftmost edge of the scale because they are the most variable languages, whereas mora-timed languages are placed at the right edge because they are invariable languages. Ramus et al. (1999) took a similar approach by considering the durational ratios between consonants and vowels and comparing ten languages. They showed that Dauer's idea was basically correct. This hypothesis seems to be convincing, but there are some problems.

First, in so far that the complexity of syllable structure and the presence of vowel reduction show differences between stress- and syllable-timing, Ramus et al.'s findings sound reasonable enough to maintain an argument for the distinction between these two types of rhythm timings. However, it seems that the degrees of complexity of syllables in syllable-timed and mora-timed languages cannot be differentiated with this approach since, as pointed out in section 2.1, these two rhythmic types of languages give similar auditory impressions.

Second, if the argument proposed by Ramus et al. is correct, any language whose syllable structure is simple enough may be regarded as a mora-timed language. For example, according to Blevins (1995), Hausa is a language whose syllable structure is basically CV. However, this language has not been reported as a mora-timed language. This fact seems to suggest that phonological factors alone may not determine mora-timing.

6.4 Acquisition of mora-timing

It is probably worth discussing mora-timing in relation to language acquisition (see Ota, this volume, for full discussion of phonological acquisition in Japanese). The

study of speech rhythm was originally investigated for the purpose of description in the field of phonetics (Pike 1945; Abercrombie 1967). However, since 1980s, this interest has been directed towards the acquisition of spoken languages in the field of psycholinguistics. This was mainly motivated by the idea that newborn babies may utilize speech rhythm to construct the lexicon because they acquire their native languages by being simply exposed speech signals (see Mehler et al. 1981; Cutler et al. 1986; see also Cutler 2012 for a comprehensive review).

In order to prove this hypothesis for mora-timing, it is vital to demonstrate which recurring unit plays a significant role. As we have seen, we have proposed three kinds of recurring units. The question is which unit is most plausible for newborn babies to perceive. Needless to say, they have no knowledge of a lexicon, so that no phonological information would be available to them when they cope with the incoming speech signals. If mora-timing is determined by the notion of mora, it would probably be difficult for babies to perceive speech signals because the notion of mora has to be acquired. Interestingly, recent studies on Japanese babies report that babies are sensitive to the distinction in vowel length (long vs. short vowels) and consonant length (geminate vs. singleton consonants) (Sato et al. 2010; Sato et al. 2012). However, even if they can make the distinction in vowel and consonant length, it does not necessarily mean that they can understand the notion of mora. Since the Japanese lexicon is structured by mora boundaries (Cutler and Otake 2002), Japanese babies must acquire this knowledge. Thus, the recurring unit on mora-timing alone proposed by Ramus et al. may not be workable for newborn babies.

We should ask then when and how babies acquire the notion of mora. There are many studies on phonological awareness of mora by Japanese children (e.g., Mann 1986). However, these studies are problematic because, as we saw, *kana* letters and moras largely correspond to each other. One possible solution may be to examine the ways children can extract words that are embedded in longer words (see Cutler 2012 for a general review). It is now well recognized that words are embedded within other words. If children without the knowledge of the writing system can extract embedded words, this strategy may tell us what boundary knowledge they have. A recent study on spontaneous puns shows that embedded words are used as puns in Japanese (see Otake and Cutler 2013). If children can manage to do the same thing, that may provide us with new evidence.

7 Conclusion and future issues

In this chapter, we reviewed mora and mora-timing from various viewpoints. First, we showed that the notion of mora was interpreted in different ways during the three periods of linguistic research. Different interpretations emerged depending how the properties of mora were treated. During the period of classical languages,

both durational and structural properties of mora were exploited to explain linguistic phenomena. The durational property was used to explain the concrete linguistic activity (pronunciation), whereas the structural property was used to explain abstract linguistic structures (the organization of rhythm and the location of accent). In the structural linguistics period, the structural property of mora was mainly emphasized to explain the ways the prosodic properties were assigned to mora. The durational property, on the other hand, was arbitrarily used without referring to any particular linguistic phenomenon. In more recent years, the structural property was mainly used to explain various linguistic phenomena.

Second, we demonstrated how the properties of mora were exploited for mora-timing. The mora-timing hypothesis focuses on the durational property of mora, whereas the mora boundary hypothesis looks primarily at the structural property. However, the more recent version of mora-timing is concerned with the degree of variability of syllables without reference to the notion of mora.

Given their merits and demerits, all these proposals make mora-timing a more complicated issue now. It is important to note that while mora is an integral part of the weight system, the Japanese syllable system in fact includes an underlying device which splits heavy syllables into two independent units. It seems that this is closely related to the notion of mora-timing. However, researchers tend to look only at Japanese in the investigations on this type of speech rhythm. It is definitely necessary to look at more languages that are called mora-timed languages in order to provide a full picture.

Acknowledgments

The earlier version of the present paper was prepared at Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, during my visit in July and September, 2012. I would like to thank to Anne Cutler for inviting me as a visiting researcher. I would also like to thank Donna Erickson and Haruo Kubozono for valuable comments on both the contents and the style of this paper. All errors are of course mine.

References

Abercrombie, David. 1967. *Elements of general phonetics*. Chicago: Aldine.
Allen, W. Sidney. 1973. *Accent and rhythm*. Cambridge: Cambridge University Press.
Balasubramanian, Tyagaraja. 1980. Timing in Tamil. *Journal of Phonetics* 8. 449–467.
Balasubramanian, Tyagaraja. 1981. Duration of vowels in Tamil. *Journal of Phonetics* 9. 151–161.
Beckman, Mary E. 1982. Segment duration and the 'mora' in Japanese. *Phonetica* 39. 113–135.
Bekku, Sadanori. 1977. *Nihongo no rizumu* [Rhythm in Japanese]. Tokyo: Kodansha.
Bingham, William. 1871. *A grammar of Latin language: For the use of schools and colleges, with exercises and vocabularies*. Philadelphia: E.H. Butler.
Blevins, Juliette. 1995. The syllable in phonological theory. In John Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Cambridge, MA: Blackwell.
Bloch, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26. 86–125.

Bosch, Anna R. K. 2011. Syllable-internal structure. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 781–798. Malden, MA and Oxford: Wiley-Blackwell.
Broselow, Ellen. 1995. Skeletal positions and moras. In John Goldsmith (ed.), The *handbook of phonological theory*, 175–205. Cambridge, MA: Blackwell.
Chamberlain, Basil Hall. 1887. *The classical poetry of the Japanese*. London: Trubner Press.
Cutler, Anne. 2012. *Native listening*. Cambridge, MA: MIT Press.
Cutler, Anne, Jacques Mehler, Dennis Norris and Juan Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25. 385–400.
Cutler, Anne and Takashi Otake. 1994. Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33. 824–844.
Cutler, Anne and Takashi Otake. 2002. Rhythmic categories in spoken-word recognition. *Journal of Memory and Language* 46(2). 296–322.
Dasher, Richard and Dwight L. Bolinger. 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12. 58–69.
Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51–62.
Donaldson, William John. 1848. *A complete Greek grammar for the use of learners*. London: Parker.
Halporn, James W., Martin Ostwald and Thomas G. Rosenmeyer. 1963. *The Meters of Greek and Latin poetry*, revised edition. London: Methuen.
Han, Mieko S. 1962. The feature of duration in Japanese. *The Study of Sounds* (The Phonetic Society of Japan) 10. 65–80.
Haraguchi, Shosuke. 1996. Syllable, mora and accent. In Takeshi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 45–75. Berlin: Mouton de Gruyter.
Hattori, Shiro. 1960. *Gengogaku no hōhō*. [Methodologies in linguistics]. Tokyo: Iwanami.
Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago press.
Hockett, Charles. 1955. *A manual of phonology* (Memoir 11). Bloomington: Indiana University Publications in Anthropology and Linguistics.
Homma, Yayoi. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9. 273–281.
Hyman, Larry. 1985. *A theory of phonological weight*. Dordrecht: Foris.
Inagaki, Kayoko, Giyoo Hatano and Takashi Otake. 2000. The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology* 75. 70–91.
Kamei, Takashi. 1956. On'in no gainen wa nihongo ni yūyō nariya. *Kokubungakukō* 15. 1–11. Hiroshima Daigaku Kokugo Kokubun Gakkai.
Katada, Fusa. 1990. On the representation of moras: Evidence from a language game. *Linguistic Inquiry* 21(4). 641–646.
Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kawamoto, Koji. 1990. *Nihon shīka no dentō – Shichi to go no shigaku* [The tradition of Japanese verse – Study of poetry with 7 and 5]. Tokyo: Iwanami.

Kindaichi, Haruhiko. 1967. Onsetsu, mōra oyobi haku [Syllable, mora and haku]. In Haruhiko Kindaichi (ed.), *Nihongo on'inron no kenkyū*, 58–77. Tokyo: Tōkyōdō Shuppan.

Klatt, Denis H. 1976. The linguistic use of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59. 1208–1221.

Komatsu, Hideo. 1977. *Nihongo no sekai 7: Nihongo no on'in* [The world of Japanese 7: Japanese phonology]. Tokyo: Chūōkōronsha.

Kubozono, Haruo. 1995. Perceptual evidence for the mora in Japanese. In Bruce Connell and Amalia Arvaniti (eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 141–156. Cambridge: Cambridge University Press.

Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lloyd, James Arthur. 1940. *Speech signals in telephony.* London: Pitman & Sons.

Low, Ee Ling, Esther Grabe and Francis Nolan. 2000. Quantitative characterization of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43. 377–401.

Mann, Verginia A. 1986. Phonological awareness: The role reading experience. *Cognition* 24. 65–92.

Marty, Larita, Takashi Otake and Anne Cutler. 2007. Perception tests of rhythmic similarity: I. Mora rhythm. *Language and Speech* 50. 77–99.

McCarthy, John J. and Prince Alan. 1986. Prosodic morphology. Unpublished ms., University of Massachusetts, Amherst and Brandeis University.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCawley, James D. 1978. What is a tone language? In Victoria Fromkin (ed.), *Tone: A linguistic survey*, 113–131. New York: Academic Press.

Mehler, Jacques, Jean Yves Dommergues, Uli Frauenfelder and Juan Segui. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Behavior* 20. 298–305.

Moseley, William. 1827. *A dictionary of Latin quantities: Or prosodican's guide to the different quantities of every syllable in the Latin language, alphabetically arranged*. London: n.p.

Nespor, Marina, Mohinish Shukla and Jacques Mehler. 2011. Stress-timed vs. syllable-timed languages. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1140–1159. Malden, MA and Oxford: Wiley-Blackwell.

Odden, David. 2011. The representation of vowel length. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1335–1361. Malden, MA and Oxford: Wiley-Blackwell.

Okimori, Takuya. 2010. *Nihongo no rekishi* [The history of the Japanese language]. Tokyo: Bere Shuppan.

Okumura, Mitsuo. 1972. Kodai no on'in [Phonology of Ancient Japanese]. In Norio Nakata (ed.), *On'inshi to mojishi* [History of phonology and orthography], 63–171. Tokyo: Taishukan.

Ota, Mitsuhiko. this volume. Chapter 17: L1 phonology: phonological development. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Otake, Takashi. 1990. Rhythmic structure of Japanese and syllable structure. *Technical Report* No. 89, 55–61. Tokyo: Institute of Electronics, Information and Communication Engineers.

Otake, Takashi and Anne Cutler. 2013. Lexical selection in action: Evidence from spontaneous punning. *Language and Speech* 56(4). 554–572.

Otake, Takashi, Giyoo Hatano and Kiyoko Yoneyama. 1996. Speech segmentation by Japanese listeners. In Takashi Otake and Anne Cutler (eds.), *Phonoloical structure and language processing: Cross-linguistic studies*, 183–201. Berlin: Mouton de Gruyter.

Otake, Takashi, Giyoo Hatano, Anne Cutler and Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32. 258–278.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Pike, Kenneth L. 1947. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: Michigan University Press.

Port, Robert, Al-ami Salman and Shosaku Maeda. 1980. Temporal compensation and universal phonetics. *Phonetica* 37. 235–252

Port, Robert, Jonathan F. Dalby and Michael O'Dell. 1987. Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America* 81(5). 1574–1585.

Ramus, Franck, Marina Nespor and Jacques Mehler. 1999. Correlates of linguistic rhythm in speech signal. *Cognition* 73. 265–292.

Sato, Yutaka, Yuko Sogabe and Reiko Mazuka. 2010. Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology* 46(1). 106–119.

Sato, Yutaka, Mahoko Kato and Reiko Mazuka. 2012. Development of single/geminate obstruent discrimination by Japanese infants: Early integration of durational and nondurational cues. *Developmental Psychology* 48(1). 18–34.

Trager, George L. 1941. The theory of accentual system. In Leslie Spier, A. Irving Hallowell and Stanley S. Newman (eds.), *Language, culture and personality: Ess*ays in memory of Edward Sapir, 131–145. Menasha, Wisconsin: Sapir Memorial Publication Fund.

Trubetzkoy, Nikolai. 1969. *Principles of phonology.* Berkeley: University of California Press.

Vance, Timothy. 1987. *An introduction to Japanese phonology.* Albany, NY: State University of New York Press.

Warner, Natasha and Takayuki Arai. 2000. Japanese mora-timing: A review. *Phonetica* 58. 1–25.

Zec, Draga. 2011. Quantity-sensitivity. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 2*, 1335–1361. Malden, MA and Oxford: Wiley-Blackwell.

Yosuke Igarashi

13 Intonation

1 Introduction

1.1 Background and the aims of this chapter

Despite a handful of pioneering frameworks (e.g., Kawakami 1957b; Fujisaki and Sudo 1971), the intonation system of (Standard) Tokyo Japanese remained one of the most understudied areas in Japanese phonology until the 1980s, not least because only limited instrumental techniques to characterize the intonation contours of utterances were available. However, the generally rudimentary nature of the study of intonation should be attributed primarily to the absence of any widely accepted framework for the description of intonational phenomena not only in Japanese but also in other languages (Ladd 1996, 2008).

Bruce's (1977) study on Swedish word accents and Pierrehumbert's (1980) study of English intonation achieved a breakthrough in the research on intonational phonology, and they provided a common framework for the description of the intonation systems of a variety of languages. This framework, generally referred to as the Autosegmental-Metrical (AM) model of intonational phonology, was successfully applied to Japanese (Pierrehumbert and Beckman 1988), and several findings related to this language contributed to the development of the general AM theory (Ladd 1996, 2008; Gussenhoven 2004; Jun 2005).

This chapter provides an overview of the intonation structure of Japanese primarily based on the X-JToBI framework (Maekawa et al. 2002; Venditti, Maekawa, and Beckman 2008), an extended version of the original Japanese Tone and Break Indices, or J_ToBI (Venditti 2005). The Tone and Break Indices system, or ToBI, is a set of conventions for transcribing and annotating the intonation patterns and prosodic structure of languages (Silverman et al. 1992). The ToBI system was firstly proposed for English but has been applied to a number of languages such as German, Greek, Korean, and Japanese, and its effectiveness has been confirmed by a growing body of studies (see Jun 2005). Both X-JToBI and J_ToBI owe their theoretical foundation to the major study of the Japanese intonation system by Pierrehumbert and Beckman (1988), which is based on the AM theory of intonational phonology. Among the existing descriptive frameworks X-JToBI can be regarded as the most useful one for the aim of this chapter, since this model was applied to a large scale database of Japanese spontaneous speech (Maekawa 2003) and underwent a number of improvements so as to describe a wide range of intonational phenomena that are frequently observed in spontaneous speech but are difficult to elicit in lab speech (such as various types of boundary pitch movements (see section 3)).

Sections 2 and 3 describe two major elements in the Japanese intonation system: prosodic phrasing and boundary pitch movements, respectively. Section 4 discusses the implications of Japanese intonation on the theory of intonation in general, highlighting similarities and differences between Japanese and other languages. Section 5 concludes this chapter. The rest of this section provides a background, describing prosodic phrasing and boundary pitch movements as well as laying out the approach that this chapter adopts for defining prosodic phrases.

All of the example utterances are produced by the author (a native speaker of Tokyo Japanese) as a means of illustration. Readers who are interested in spontaneous speech utterances are advised to refer to Venditti, Maekawa, and Beckman (2008), in which a number of examples are taken from a Japanese speech database. The terms "pitch" and "fundamental frequency" (F0) are used interchangeably in this chapter.

1.2 Major elements of Japanese intonation

The Japanese intonation system can be described in terms of two major elements: *prosodic phrasing* above the word level (prosodic phrasing, henceforth) and *boundary pitch movements*. Since prosodic phrasing in Japanese is closely connected with lexical pitch accent, we will also discuss this in some detail.

Japanese has two types of words in its lexicon generally referred to as "accented words" and "unaccented words" (Kawahara, this volume). The former exhibit a pitch contour with a steep fall from high (H) to low (L) somewhere in the word, while the latter show a contour with no such fall. The term *pitch accent* in this chapter refers to this lexically specified pitch fall in the accented words. For example, *a'me* 'rain' has a pitch accent in the initial syllable and is therefore an accented word, whereas *ame* 'candy' has no pitch accent and is therefore an unaccented word. (The accented vowel is post-marked by an apostrophe.) In addition to the presence/absence of pitch accent, its location in the word is also lexically specified in Japanese. For example, *na'mida* 'tear' has an accent in the initial syllable, *nomi'ya* 'bar' in the second, and *atama'* 'head' in the final.

Prosodic phrasing is defined as the grouping of linguistic units (often words) in an utterance by means of prosodic, or suprasegmental features such as pitch, intensity, and duration. In Japanese, prosodic phrasing is achieved primarily by the modification of the pitch patterns of the utterance. Prosodic phrasing can signal the focused word in the sentence, the syntactic constituency of the sentence, and so forth (see Ishihara, this volume).

It is generally accepted that Japanese has two levels of prosodic phrasing (McCawley 1968; Poser 1984; Kubozono 1988/1993; Maekawa et al. 2002; Venditti 2005). This chapter also adopts this "double-layered model" of prosodic phrasing, although both levels of phrasing are not without some controversy.

Boundary pitch movements are tones that occur at the end of the prosodic constituent in Japanese and contribute to the pragmatic interpretation of the utterance,

such as questioning, continuation, and emphasis (Venditti, Maekawa, and Beckman 2008). For example, when the sentence *ta'roo-ga kuruma-o kat-ta* (taroo-NOM car-ACC buy-PAST 'Taro bought a car.') is produced with a rising pitch at the utterance-final mora, the sentence is interpreted as a question, whereas the same sentence without such a boundary pitch movement is interpreted as a statement. BPM is sometimes referred to as "sentence-final intonation" (e.g., Vance 2008). This terminology is avoided in this chapter, since BPM not only occurs sentence-finally, but also sentence-medially.

1.3 Syntactic vs. intonational phonological approaches

There is ample evidence to show that, in general, prosodic phrasing cannot be predicted from syntax alone (Bolinger 1972). There can be multiple options in the phrasing of utterances with the same syntax. It is also well known that certain extra-syntactic factors such as speech rate and length of utterance play a role in determining prosodic phrasing (see Shattuck-Hufnagel and Turk 1996 for a review). There is currently intense debate about how syntax and prosody are related to each other (e.g., Nespor and Vogel 1986; Truckenbrodt 1999; Selkirk 2000, 2009; Ito and Mester 2012), but as yet no consensus has emerged. Consequently, there is no general agreement on how prosodic phrases are defined. To put it briefly, opinions differ depending on to what extent prosodic phrasing is predicted from syntactic structure. This chapter adopts what Jun (1998) calls the "intonational phonological approach" as opposed to the "syntactic approach". In the former approach, a prosodic phrase is defined on the basis of the surface phonetic form of the utterance, specifically its pitch contour, without reference to its syntactic structure. For explanation of the syntactic approach, see Ishihara (this volume).

2 Prosodic phrasing

2.1 Levels of prosodic phrasing

In the X-JToBI scheme, prosodic phrasing occurs at two levels. The lower level is the Accentual Phrase, while the higher level is the Intonation Phrase. The two prosodic domains are hierarchically organized and obey the Strict Layer Hypothesis (Selkirk 1986), that is, that any domain at a given level of the hierarchy consists exclusively of domains at the next lower level of the hierarchy.

Although some of the frameworks of the Japanese intonation system posit only a single level of prosodic phrasing (Kawakami 1957b; Uwano 1989, 1999),[1] most

1 The frameworks proposed by Kawakami (1957b) and Uwano (1989, 1999) posit a single prosodic phrase called "tonal phrase" or simply "phrase", which can contain more than one accented word. This prosodic constituent is tonally marked by a delimitative pitch rise as well as a pitch range expansion at its beginning.

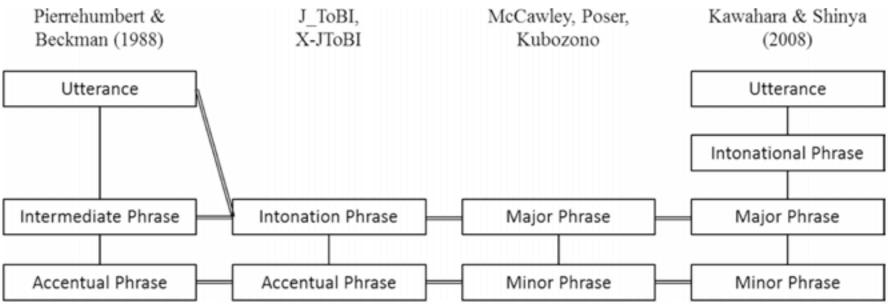


Figure 1: Terminology used by different groups of researchers for the levels of prosodic phrasing. The terminology adopted in the recursive model (e.g., Ito and Mester 2012) is not represented.

researchers adopt the double-layered model. However, not only does their terminology vary, but the definition of the prosodic phrases also slightly differs among the researchers who adopt the double-layered model. Figure 1 is a schematic illustration of the correspondence in the levels of prosodic phrasing proposed in major studies.

McCawley (1968), Poser (1984), and Kubozono (1993) refer to the lower prosodic phrase as the Minor Phrase and to the higher as the Major Phrase. Pierrehumbert and Beckman (1988), on the other hand, term them the Accentual Phrase and the Intermediate Phrase, respectively. They also posit a prosodic constituent above the Intermediate Phrase in their prosodic hierarchy and term it the utterance. Maekawa et al. (2002) and Venditti (2005) merge Pierrehumbert and Beckman's Intermediate Phrase and utterance and call it the Intonation Phrase.

On the other hand, Kawahara and Shinya (2008), following Selkirk's (1986) position, assume three levels of prosodic phrasing above the word and below the utterance: namely, Minor Phrase, Major Phrase, and Intonational Phrase. The first two prosodic phrases broadly correspond to those proposed in the double-layered models (McCawley 1968; Poser 1984; Kubozono 1988/1993), whereas their Intonational Phrase located between the Major Phrase and utterance is not posited in other frameworks. (Note that their Intonational Phrase differs from the Intonation Phrase in X-JToBI.) Kawahara and Shinya (2008) investigated the effect of syntactic gapping and coordination on Japanese intonation and argue that each syntactic clause projects its own Intonation Phrase, while an entire sentence constitutes one Utterance. Further research is needed to confirm whether this level of prosodic phrasing exists in Japanese.

A prosodic hierarchy that does not obey the Strict Layer Hypothesis has also been proposed (Ladd 1986, 1988; Gussenhoven 1991; Selkirk 1996, 2000) and it has been applied to Japanese (Selkirk 2009; Ito and Mester 2012). In this framework, prosodic structure in principle allows an unlimited number of prosodic recursions of the same level of the prosodic hierarchy. Thus, Major Phrase, for example, can dominate another Major Phrase in this model. For further discussion of the recursive model,

see Ishihara (this volume). Following the X-JToBI framework, the forthcoming discussion is based on the double-layered model that obeys the Strict Layer Hypothesis.

2.2 Accentual phrase

In X-JToBI, an Accentual Phrase (AP) is defined 1) as having a delimitative rise to high around the second mora and a subsequent gradual fall to low at the end of the phrase, and 2) as having at most one lexical pitch accent. While a typical AP consists of one lexical word plus any following particles or postpositions (e.g., *yama'-ga* mountain-NOM, *niwa-ni'-wa* garden-LOC-TOP), it is not always the case. It is quite often the case that a single AP can contain two or more lexical words (e.g., *hirosima-no omiyage* Hiroshima-GEN souvenir 'a souvenir from Hiroshima'). Moreover, as we see in section 2.6, a particle can form its own AP (Okumura 1956; Sagisaka and Sato 1983; Kubozono 1988/1993; Maekawa and Igarashi 2007; Vance 2008).

In X-JToBI, the intonation contours are analyzed as a linear sequence of level tones, high (H) and low (L), irrespective of the source of tones (either lexical or post-lexical). This sequential tone structure is characteristic of the AM theory of intonational phonology and is opposed to the superposition model proposed by Fujisaki and his colleagues (Fujisaki and Sudo 1971; Fujisaki and Hirose 1984; Fujisaki 1989), in which the local word accent properties are overlaid onto the global phrasal properties. (For details of the superposition model, see Ishihara, this volume).

Thus, in X-JToBI, the intonation contours of a single AP are described as a sequence of tones that are shown below for, respectively, an unaccented AP (i.e., an AP containing no accented words (1a)), and for an accented AP (i.e., an AP containing an accented word (1b)).

(1) AP tones

a.	Unaccented AP:	%L H-		L%
b.	Accented AP:	%L H-	H*+L	L%

"H*+L" stands for pitch accent, where an asterisk indicates the tone associated with the mora that is governed by the accented syllable.[2] Henceforth, the mora with which the H tone of H*+L is associated will be referred to as the accented mora.

The L tones with a diacritic "%" are called the *boundary tones*, with %L being the initial boundary tone, and L% the final boundary tones.[3] The low tone found at the beginning of the AP is sometimes called the *initial lowering* in some frameworks (e.g., Haraguchi 1977; Selkirk and Tateishi 1991).

2 Pitch accent is transcribed as simply "HL" in Pierrehumbert and Beckman's (1988) framework, but it is tagged as "H*+L" in the original J_ToBI (Venditti 2005). The pitch accent label in X-JToBI is actually "A", while in this chapter we use "H*+L" following the original ToBI convention.

3 In the original research, Pierrehumbert and Beckman's (1988) initial and final boundary tones are transcribed as "L%", while the Japanese ToBI convention uses "%L" (with the diacritics to the left of "L") for indicating the initial boundary tone. Note that the usage of a diacritic "%" slightly differs from that in most of the ToBI systems in other languages, such as English (Beckman, Hirschberg,

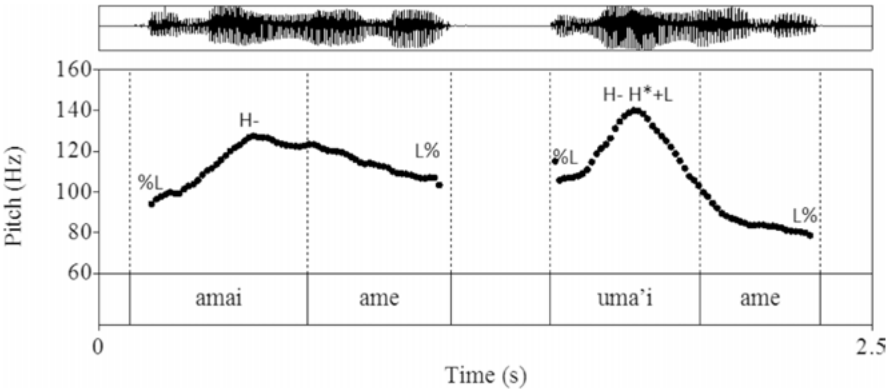


Figure 2: An unaccented AP *amai ame* 'sweet candy' (left), and an accented AP *uma'i ame* 'good-tasting candy' (right).

The H tone with a hyphen to the right is called the *phrasal high*. %L and H- function as a starting and ending point, respectively, of the AP initial rise (see below). H*+L is a pitch accent being manifested as a sharp fall in pitch only in an accented AP. L% serves as an end point of a gradual fall in pitch from H- in the case of an unaccented AP, or from L of H*+L in the case of an accented AP. In Figure 2 (left), the unaccented adjective *amai* 'sweet' is combined with the following unaccented noun *ame* 'candy' into a single unaccented AP, where the %L H- L% pattern can be clearly observed.[4]

A delimitative rise observed at the beginning of the AP is deemed in the AM framework as a sequence of a boundary tone and a phrasal high. To be more specific, when the AP appears utterance-initially or follows a pause, a string of the initial boundary tone (%L) and the phrasal high (H-) is manifested as the initial rise. When two or more APs are concatenated without a pause intervening between them, a string of the AP-final boundary tone (L%) and the phrasal high (H-) is

and Shattuck-Hufnagel 2005) and German (Grice, Baumann, and Benzmüller 2005). In the latter, the diacritic "%" signifies not only that the tone is a boundary tone, but also that the tone is linked to the same prosodic constituent, specifically, the intonation phrase. Similarly, the diacritic "-" stands for the boundary tone that is linked to the prosodic constituent that is lower than the intonation phrase, namely, the intermediate phrase. In X-JToBI, by contrast, the difference in the diacritics does not indicate that tones are liked with the different prosodic constituents. For example, H- and L% are, although the diacritics differ, linked to the same prosodic constituent, that is, the AP. Incidentally, readers should not confuse "boundary tone" here with "boundary pitch movements" discussed in section 3 below.

4 In the X-JToBI labeling convention, H- of an accented AP is labeled only when the peak of the phrasal high is distinguishable from the peak of the lexical pitch accent. This is merely a practical treatment that minimizes the labeling cost. In Pierrehumbert and Beckman's (1988) model, the phrasal high (H-) is never delinked from the left edge of the AP. Following this model, H- is not omitted in the example utterances in this chapter, even when the corresponding F0 event is not observed.

realized as the initial rise of the non-initial APs. In other words, so-called Initial Lowering is considered in X-JToBI as a manifestation of either %L or L%, depending on the left context of the AP.

Readers familiar with traditional studies on Japanese prosody (e.g., Kawakami 1957b; Uwano 1989, 1999) might be confused with this rather cryptic treatment of the initial rise. In such studies, the initial rise is a property of the beginning of the AP. In the AM framework, by contrast, the rise is decomposed into L and H, and the L that serves as the beginning of the rise changes the affiliation to the prosodic units (either %L or L%) depending on context. In fact, the status of the first %L tone is somewhat unclear in the AM frameworks of Japanese intonation. In Pierrehumbert and Beckman's (1988) model, it is a tone associated with the left edge of the utterance. In the J_ToBI framework, in which the utterance and intermediate phrase are merged into a single prosodic phrase (i.e., Intonation Phrase), %L is a tone that appears when the AP follows a pause (Venditti 2005), and the affiliation of this tone is not made explicit.[5]

2.3 Association of the AP tones

Although they are not fully discussed from the guidelines of both J_ToBI (Venditti 2005) and X-JToBI (Maekawa et al. 2002), complex tone association rules are proposed in Pierrehumbert and Beckman's (1988) model in order to account for surface pitch patterning of the APs. First, the H of the accentual H*+L is linked to the lexically accented mora with the L being unlinked. In other words, the H and L are regarded as an indivisible unit at some level of analysis (for full discussion, see Chapter 5.4 of Pierrehumbert and Beckman's 1988 book). The association of accentual H*+L tones are shown in (2).

(2) Linking of accentual tones:

a. (nimono)
'boiled food'

b. (na'mida)
H*+L
'tear'

c. (nomi'ya)
H*+L
'bar'

d. (atama')
H*+L
'heard'

5 In the practical guideline of X-JToBI (Igarashi et al. 2006), which is proposed for the labeling of the Corpus of Spontaneous Japanese (CSJ, see section 3.7), the beginning of the AP is always marked by %L. Although it is not explicitly stated in the guideline, this implies that %L belongs to the AP. In the X-JToBI labels provided in CSJ, therefore, the boundary between two successive APs is tagged by two labels, L% and %L.

The unaccented AP (2a) has no pitch accent. In the accented APs, H of H*+L is associated with the first (2b), second (2b) and final (2d) moras, respectively.

It is a common observation that the initial rise of the AP is almost imperceptible when 1) the AP-initial mora is lexically accented, and/or 2) the AP-initial syllable is heavy and sonorant, i.e., a long vowel, a diphthong, and a short vowel followed by a nasal (e.g., Hattori 1954; Haraguchi 1977). An F0 contour in these cases exhibits a small rise with the beginning of the rise being significantly higher than that seen in ordinary cases. Pierrehumbert and Beckman (1988) account for this weakly realized low tone by assuming that the initial %L (or L% in the case of the non-initial AP) is not linked directly to the minimal tone bearing unit (TBU) in the prosodic hierarchy (i.e., mora in the case of Japanese). This is exemplified in (3).

(3) Association of %L

a. (nomi'ya)
%L H*+L
'bar'

b. (na'mida)
%L H*+L
'tear'

c. (yoosi)
%L
'paper'

d. (saihu)
%L
'wallet'

e. (tansu)
%L
'drawer'

%L is left unlinked with any mora when AP initial mora is lexically accented (3b), or when the AP-initial syllable is a long vowel (3c), a diphthong (3d), or a short vowel followed by a nasal (3e).

The phrasal high (H-) is linked with the second mora of the AP unless the initial or second mora is lexically accented, as shown in (4)

(4) Association of L% and H-

a. (nomi'ya)
%L H- H*+L

b. (na'mida)
%L H- H*+L

c. (yoosi)
%L H-

d. (saihu)
%L H-

e. (tansu)
%L H-

Regardless of whether they are linked with moras or not, %L and L% are linked to the edges of higher domains in the prosodic hierarchy. In Pierrehumbert and Beckman's (1988) framework, where the utterance is posited as the highest prosodic

domain in the hierarchy, %L is linked to the left edge of the utterance. L% is associated with the right edge of the AP. This tone is left unlinked with any mora except in those specific cases where another AP immediately follows. In such cases, linking of L% with the mora is governed by the same principles as for %L.

The linking of the boundary tones with the edges of prosodic phrases is sometimes called *primary association*, which contrasts with *secondary association* in which the tones are linked with the moras (Pierrehumbert and Beckman 1988; see also Grice 1995). In the same way, H- is primarily associated with the left edge of the AP, and secondarily associated with the second mora (unless the first or second mora is accented). The primary and secondary associations of %L, H-, and L% are shown in (5), where "{ }" represents the boundaries of the utterance. Primary and secondary association is indicated by dashed and solid lines, respectively. The level of the intermediate phrase is omitted.

(5) Primary and secondary associations of %L, H-, and L%

a. { (nomi'ya) }
%L H- H*+L L%

b. { (na'mida) }
%L H- H*+L L%

c. { (yoosi) }
%L H- L%

d. { (saihu) }
%L H- L%

e. { (tansu) }
%L H- L%

In X-JToBI, where the utterance is not posited in the prosodic hierarchy, association of %L with a higher domain than the mora is not assumed, and therefore, as already mentioned in section 2.2, the affiliation of %L is not clear. For the sake of simplicity, in what follows the primary association of tones is not expressed.

2.4 Sequences of APs

When a speaker produces the fluent utterances of the sentences in (6), he groups the words in each utterance into several APs. The syntactic branching structure is [A [N_1 N_2]], where A is an adjective, and N is a noun followed by a particle, but not [[A N_1] N_2]. Some examples of prosodic phrasing of these sentences are shown in Figure 3.[6]

6 For a brief and specific description of intonation contours for combinations of unaccented and accented APs, see also Vance (2008, section 7.6).

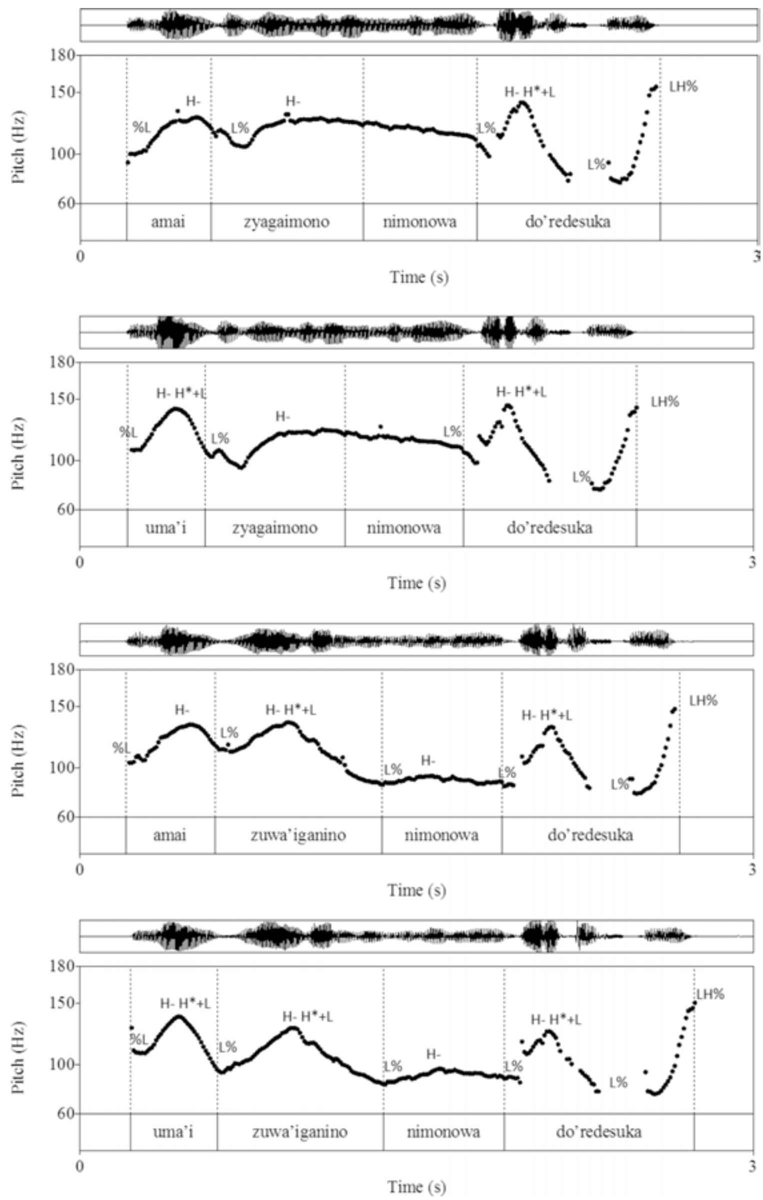


Figure 3: Phrasing at the AP level. *Amai zyagaimo-no nimono-wa do're-desu-ka?* sweet potato-NOM boiled;food-TOP which-COP-Q 'Which are the sweet boiled potatoes?' (top), *Uma'i zyagaimo-no nimono-wa do're-desu-ka?* Good-tasting potato-NOM boiled;food-TOP which-COP-Q 'Which are the good-tasting boiled potatoes?' (second from the top), *Amai zuwa'igani-no nimono-wa do're-desu-ka?* Sweet snow.crab-NOM boiled;food-TOP which-COP-Q 'Which is the sweet boiled snow crab?' (third from the top), and *Uma'i zuwa'igani-no nimono-wa do're-desu-ka?* Good-tasting snow.crab-NOM boiled;food-TOP which-COP-Q 'Which is the good-tasting boiled snow crab?' (bottom).

(6) a. amai zyagaimo-no nimono-wa do're-desu-ka?
sweet potato-GEN boiled.food-TOP which-COP-Q
'Which are the sweet boiled potatoes?'

b. uma'i zyagaimo-no nimono-wa do're-desu-ka?
Good-tasting potato-GEN boiled.food-TOP which-COP-Q
'Which are the good-tasting boiled potatoes?'

c. amai zuwa'igani-no nimono-wa do're-desu-ka?
sweet snow.crab-GEN boiled.food-TOP which-COP-Q
'Which is the sweet boiled snow crab?'

d. uma'i zuwa'igani-no nimono-wa do're-desu-ka?
Good-tasting snow.crab-GEN boiled.food-TOP which-COP-Q
'Which is the good-tasting boiled snow crab?'

When there is a right-branching syntactic boundary, an AP boundary is frequently inserted there. Thus in (6), the adjective *amai* or *umai* forms a single AP. When there is no right-branching boundary intervening them, an unaccented word and a word that follows it tend to be conjoined into an AP. In (6a,b), therefore, two noun phrases *zyagaimono-no nimono-wa* are conjoined into an AP. When an accented word precedes that word, the following word often forms its own AP (Vance 2008; Ito and Mester 2013), even if there is no right-branching boundary. Thus, in (6c,d), two noun phrases *zuwa'igani-no nimono-wa* form separate APs. In all the examples, the verb phrase *do're-desu-ka* constitutes a single AP. Prosodic phrasing and the linking of AP tones in these utterances are shown in (7). The tone found at the end of these utterances, that is, LH%, is a boundary pitch movement, which will be discussed in Section 3.

(7) Prosodic phrasing and the linking of tones in the utterances (6)

a. (amai) (zyagaimo-no nimono-wa) (do're-desu-ka)
%L H- L% H- L% H- H*+L L% LH%

b. (uma'i) (zyagaimo-no nimono-wa) (do're-desu-ka)
%L H- H*+L L% H- L% H- H*+L L% LH%

c. (amai) (zuwa'igani-no nimono-wa) (do're-desu-ka)
%L H- L% H- H*+L L% H- L% H- H*+L L% LH%

d. (uma'i) (zuwa'igani-no nimono-wa) (do're-desu-ka)
%L H- H*+L L% H- H*+L L% H- L% H- H*+L L% LH%

The grouping of words into APs depends on an interaction of various factors such as the word accentuation, syntactic branching structure, focus, or discourse structure (Venditti 2005). While this chapter does not delve into a question of what factors determine prosodic phrasing, the effects of focus will be discussed in sections 2.7 and 2.8. For full discussion of the factors that affect prosodic phrasing, see Ishihara (this volume).

2.5 Surface underspecification

One of the distinctive characteristics of Pierrehumbert and Beckman's (1988) framework is that tones are sparsely represented even on the surface with respect to the number of TBUs. This becomes clear when we consider the following example (8), where two words are conjoined to form a single accentual phrase. In (8a) five moras are toneless, while in (8b) four moras are toneless.

(8) Tone linking under the surface underspecification model

a. (amai yamaimo) b. (amai oni'giri)

%L H- L% %L H- H*+L L%

'sweet yam' 'sweet rice ball'

The view of sparse specification of tones on the surface, or the surface underspecification view has also been proposed in traditional framework such as Kawakami (1957b). It contrasts with the full specification view, in which every mora receives a tone on the surface. The latter view can be found in early studies on Japanese word-level prosody (Miyata 1928; Hattori 1954), including those within a generative phonology such as McCawley (1968) and Haraguchi (1977). The full specification view is also adopted in the analysis of Japanese phrase-level prosody (Poser 1984; Kubozono 1988/1993). Under this view, various phonological rules such as the tone spreading rules play a role in accounting for the surface pitch contour. Thus, in an unaccented AP (9a), the first mora is assigned a L tone and the other moras a H tone. In an accented AP (9b), on the other hand, the first mora receives a L tone, the second, third, fourth, and fifth moras acquire a H tone, and the last two moras carry a L tone (see also Kawahara, this volume, for the full specification account for surface pitch contours).

(9) Tone linking under the full specification model

a. (amai yamaimo) b. (amai oni'giri)

L H L H L

'sweet yam' 'sweet rice ball'

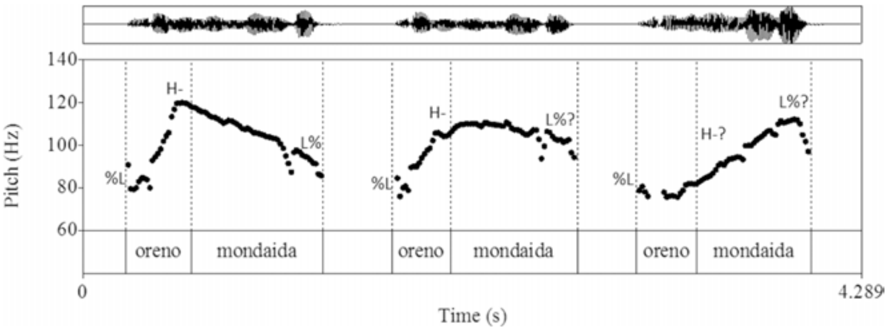


Figure 4: Variations in contour of unaccented APs *Ore-no mondai-da* I-GEN problem-COP 'It's my problem.'. An ordinal contour (left), a contour that would be accounted for by tone spreading (middle), and a rising contour without an apparent tone target (right).

The main reason for Pierrehumbert and Beckman's (1988) rejection of the full specification view is that tone spreading rules cannot account for the gradual pitch fall in the high-pitched moras (for unaccented phrases), and for that in the low-pitched moras (for accented phrases). The gradual pitch fall of interest can be observed in the second AP in the utterances in the first two panels of Figure 3. In the full specification view a possible explanation for this smooth fall is that the high-pitched moras and low-pitched moras undergo declination, that is, a non-phonological, physiological effect that lowers pitch range gradually through time regardless of what tones might be present (e.g., Fujisaki and Sudo 1971; Fujisaki and Hirose 1984; Fujisaki 1989). However, Pierrehumbert and Beckman (1988) showed that the rate of the downtrend decreases as the number of high-pitched or low-pitched moras increases, even though the declination model predicts that the rate is constant regardless of the number of moras. It is plausible that Pierrehumbert and Beckman's finding can be accounted for by postulating phonetic interpolation between H- and L% for an unaccented AP, or between L of the accentual H*+L and L% for an accented AP, with no tone spreading in either case.

On the other hand, Sugahara (2003), based on the inter-speaker variation found in her experimental results, claims that tone spreading is also possible in Japanese. The gradual pitch fall, which is most plausibly accounted for by interpolation (Figure 4 [left]), could merely be regarded as one of the possible realizations of pitch patterns in Japanese APs. In addition to a high plateau that might be a result of tone spreading (Figure 4 [middle]), a gradual rise without any turning point in the contour is also observed in an unaccented AP. Figure 4 (right) demonstrates an example in which F0 rises from the beginning of the unaccented AP to the final mora, without apparent targets for H- and L%. Although F0 rises throughout the utterance, it is interpreted as a statement, not a question. Various factors should be taken into consideration to explain the contours of APs. Variation in contour at the beginning of the AP will be discussed in section 3.6.

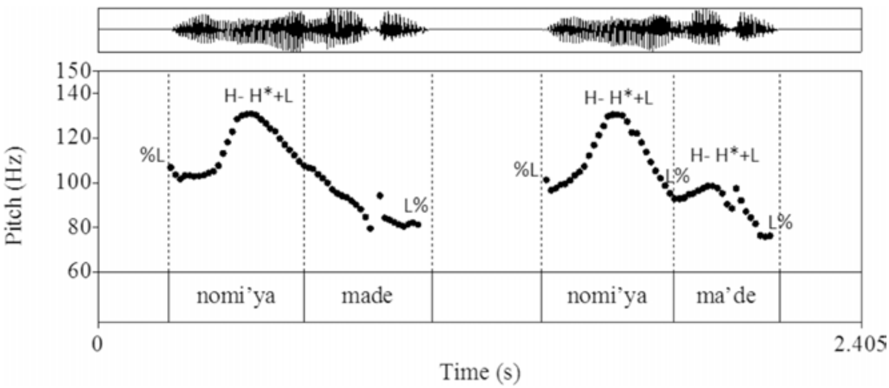


Figure 5: Deletion and retention of the pitch accent of particles. An AP with deaccenting *nomi'ya-made* 'to the bar' (left), and an AP without deaccenting *nomi'ya-ma'de* 'to the bar' (right).

2.6 Deaccenting of particles

It has been documented (e.g., Poser 1984) that lexically accented particles such as *-ma'de* 'up to', *-su'ra* 'even', and *-ko'so* 'just' lose their pitch accent when preceded by an accented lexical word, and thus the lexical word and the particle are merged into a single accented AP (e.g., *nomi'ya* 'bar' + *-ma'de* 'up to' → *nomi'ya-made* 'to the bar'). Deaccenting of particles is, however, not obligatory (Okumura 1956; Sagisaka and Sato 1983; Kubozono 1988/1993; Maekawa and Igarashi 2007; Vance 2008). The lexical accent of particles can survive, in some cases with a large pitch range, even when preceded by an accented lexical word. This means that lexically accented particles can form their own AP. This is shown in Figure 5. Prominence in particles will be discussed in section 2.4.

This not only applies to the lexical pitch accent of particles but also to their morphologically derived accent. When a sequence of unaccented particles such as *-ni-mo* and *-kara-wa* follows a noun, pitch accent is inserted to the final mora of the preceding particles (e.g., *hirosima* 'Hiroshima' + *-ni* + *-mo* → *hirosima-ni'-mo* 'in Hiroshima, too'; *hirosima* + *-kara* + *-wa* → *hirosima-kara'-wa* 'from Hiroshima') (Akinaga 2002). These morphologically inserted accents obligatorily appear on the surface when preceded by an unaccented word. In contrast, when preceded by an accented word, morphologically inserted accents can either be deleted (e.g., *ao'mori-kara-wa* 'from Aomori') or retained (e.g., *ao'mori-kara'-wa* 'from Aomori'). In the latter case, a string of particles (such as *-kara'-wa*) constitutes its own AP.

2.7 Dephrasing caused by focus

It is proposed in Pierrehumbert and Beckman's (1988) model that focus can delete AP boundaries of post-focal APs, so that the focused AP and post-focal APs are

merged into a single AP, with the focused AP being at its left-most position. The deletion of AP boundaries is called *dephrasing*. (Note that Pierrehumbert and Beckman 1988 do not argue that focus *obligatorily* causes dephrasing. Instead, dephrased utterances are taken as one of the possible realizations of focus.)

Figure 6 illustrates the utterances that have an accented or unaccented adjective followed by an accented or unaccented noun. The utterances in the right panels have focus on the first words (*Yamada-ga* or *Ya'mano-ga*). It can be seen from the figure that, except the final utterance (at the bottom), medial L% and H- in the second words (*yaoya-ni* or *nomi'ya-ni*) are lost when they are focused, resulting in a smooth transition between H- and L% or between L of H*+L and L%.

Pierrehumbert and Beckman (1988) further argue that the contours for the two utterances with focus, *YAMADA-GA yaoya-ni ori-ma'si-ta* (right, third from the top) and *YA'MANO-GA nomi'ya-ni ori-ma'si-ta* (right, bottom) in Figure 6 are identical, meaning that the lexical pitch accent of the post-focal noun can be deleted by focus. Indeed, no sharp fall corresponding to pitch accent is observed in the contour of the post-focal accented word in Figure 5 (bottom, right). However, based on his production experiment, Maekawa (1994) showed that although no sharp fall is detected, there are systematic differences in the contour between a post-focal accented word and an unaccented word. Specifically, the line fitted to the contour of the post-focal accented word ($y = ax + b$, where y = F0, x = time, a = slope, and b = intercept) has a larger intercept value and smaller slope value than that of the post-focal unaccented word. Maekawa (1997) also revealed that the differences can be perceived by listeners. Thus, for example, an accented verb *yo'n-deru* 'read' and an unaccented verb *yon-deru* 'call' as shown in Figure 7 can be distinguished correctly in the sentences *Da're-ga yo'n-deru?* 'Who is reading?' vs. *Da're-ga yon-deru?* 'Who is calling?'. In contrast, no evidence has been reported showing post-focal deaccenting in Japanese. Therefore, it is not plausible to say that dephrasing occurs in a sequence of a focused accented AP and following post-focal accented APs, since, by definition of the AP, dephrasing accompanies the deaccenting of post-focal accented APs in such a sequence.

Moreover, Sugahara (2003), on the basis of her experimental results, argues against AP-level dephrasing not only in the case of post-focal accented APs but also in the case of post-focal *unaccented* APs. Thus, although it has been claimed that focus can delete AP boundaries, a growing body of evidence has shown that focus-induced dephrasing is rarely, if at all, observed. Further examination is required to show whether focus never causes dephrasing, especially in the condition when an unaccented AP is focused (as in the examples in the top two panels in Figure 6).

2.8 Downstep and the Intonation Phrase

Now we turn to the Intonation Phrase (IP). The IP is defined as the prosodic domain immediately above the AP in the hierarchy, within which pitch range is specified.

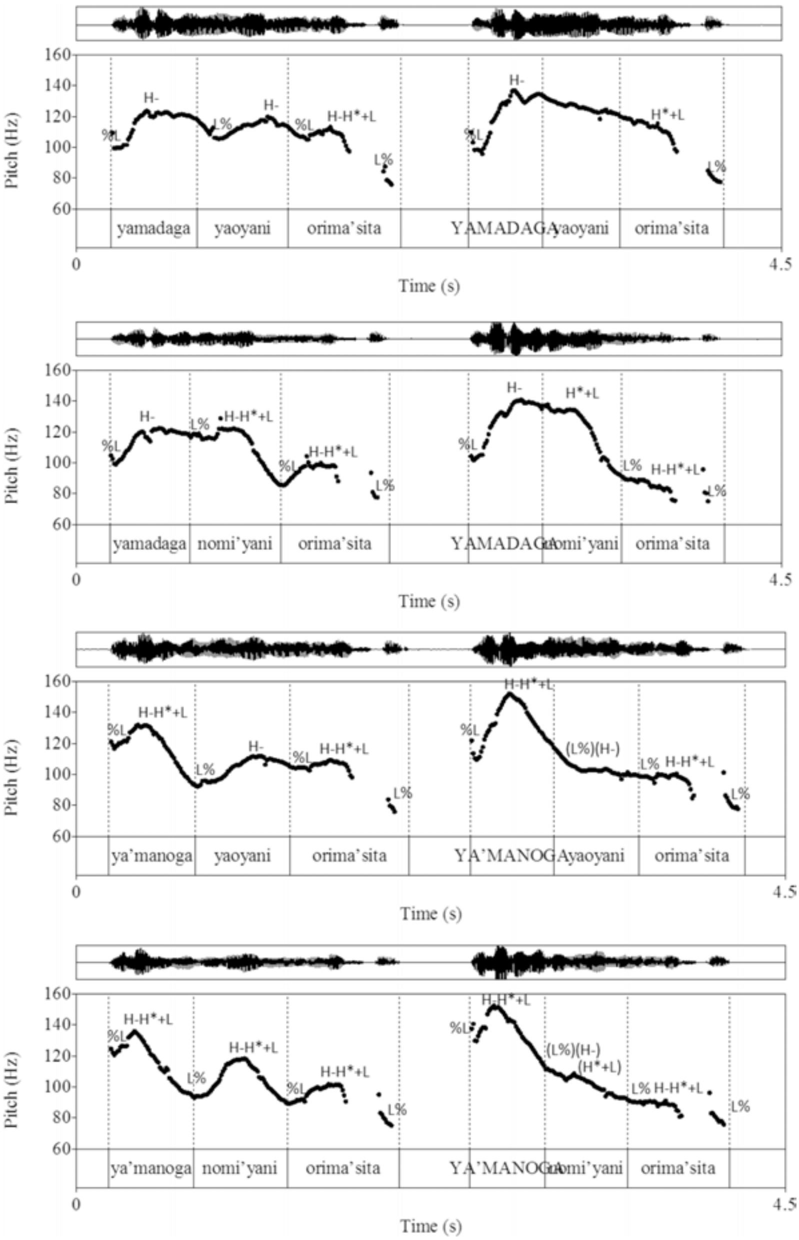


Figure 6: Dephrasing caused by focus on the adjectives. *Yamada-ga yaoya-ni ori-ma'si-ta* Yamada-NOM vegetable.shop-LOC be-POL-PAST 'Yamada is in a vegetable shop' (top), *Yamada-ga nomi'ya-ni ori-ma'si-ta* Yamada-NOM bar-LOC be-POL-PAST 'Yamada is in a bar' (second from the top), *Ya'mano-ga yaoya-ni ori-ma'si-ta* Yamano-NOM vegetable.shop-LOC be-POL-PAST 'Yamano is in a vegetable shop' (top), *Ya'mano-ga nomi'ya-ni ori-ma'si-ta* Yamano-NOM bar-LOC be-POL-PAST 'Yamada is in a bar' (second from the top). Figures in the right panels shows the utterances with focus on the first word, where the focused words are capitalized.

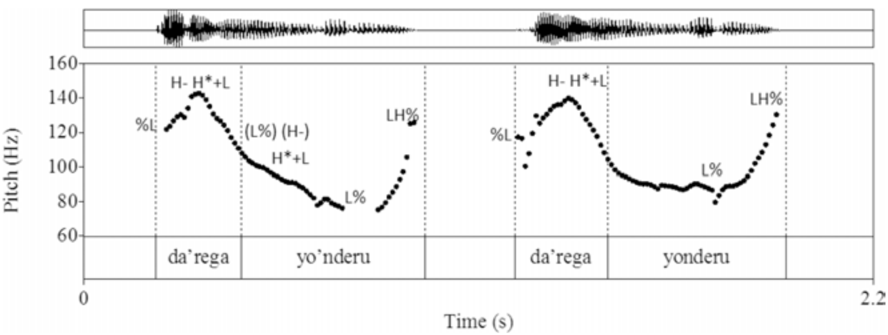


Figure 7: Distinction between post-focal accented and unaccented words. *Da're-ga yo'n-deru?* 'Who is reading?' (left) and *Da're-ga yon-deru?* 'Who is calling?' (right).

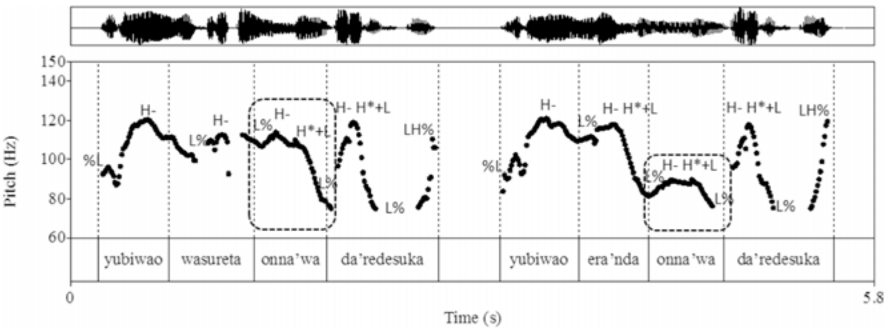


Figure 8: Downstep. An utterance without downstep *yubiwa-o wasure-ta onna'-wa dare-desu-ka?* ring-ACC forget-PAST woman-TOP who-COP.POL-Q 'Who is the woman that left the ring behind?' (left), and an utterance with downstep on the third AP *yubiwa-o era'n-da onna'-wa da're-desu-ka?* ring-ACC choose-PAST woman-TOP who-COP.POL-Q 'Who is the woman that chose the ring?' (right). The relevant portions of the F0 contours are marked by squares. Dotted vertical lines stand for AP boundaries.

At the beginning of each new IP, the speaker chooses a new pitch range, which is independent of the specification of the preceding AP (Venditti 2005). This process is called *pitch reset*. The pitch-range specification of IPs is closely connected with a phonological process called *catathesis*, or *downstep*, by which the pitch range of each AP is compressed when that AP follows an accented AP. Downstep is displayed in Figure 8. It can be seen that the peak of the third AP is significantly lower when preceded by the accented AP (right) than when preceded by the unaccented AP (left).

When multiple accented APs form a single IP, downstep occurs iteratively, and we can then observe a staircase-like F0 contour. This is demonstrated in the top panel of Figure 9. In a sequence of four APs (in a syntactic phrase with a uniformly

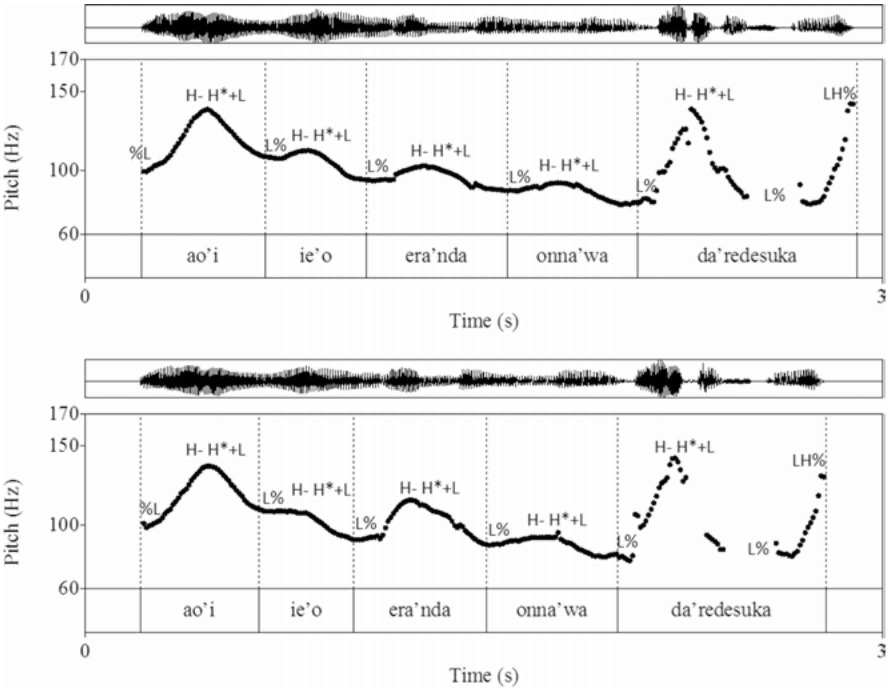


Figure 9: Successive downstep. *Ao'i ie'-o era'n-da onna'-wa da're-desu-ka?* blue house-ACC choose-PAST woman-TOP who-COP.POL-Q 'Who is the woman that chose the blue house?', without (top) and with (bottom) the rhythmic effect. Vertical lines indicate AP boundaries.

left-branching structure) as in Figure 9, however, the pitch range of the third AP is frequently expanded, so that a staircase-like F0 contour is not observed. This effect is known as "rhythmic boost" (Kubozono 1988/1993, 1989), which is often claimed to result from the Principle of Rhythmic Alternation (Selkirk 1986). The rhythmic effect is shown in the bottom panel of Figure 9, in which the pitch range of the third AP is larger than the preceding AP.

When the IP boundary is inserted, downstep is blocked at this boundary; that is, pitch reset occurs at the boundary, and a new pitch range is specified to the IP. Various linguistic factors bring about pitch reset at the IP boundary. These include syntactic constituency and focus (Kindaichi 1951; Kawakami 1957a; Uyeno, Hayashibe, and Imai 1979; Fujisaki 1989; Selkirk and Tateishi 1991; Maekawa 1994; Kori 1997; Ito 2002; Kitagawa 2005; Ishihara 2007; Kubozono 2007).

Focus has been claimed to be one of the main triggers for the insertion of an IP boundary at the beginning of the focused word (for an argument against this claim, see Ishihara 2007). Moreover, post-focal words are prosodically subordinated to the focused word. In addition to the possible dephrasing of post-focal APs mentioned in

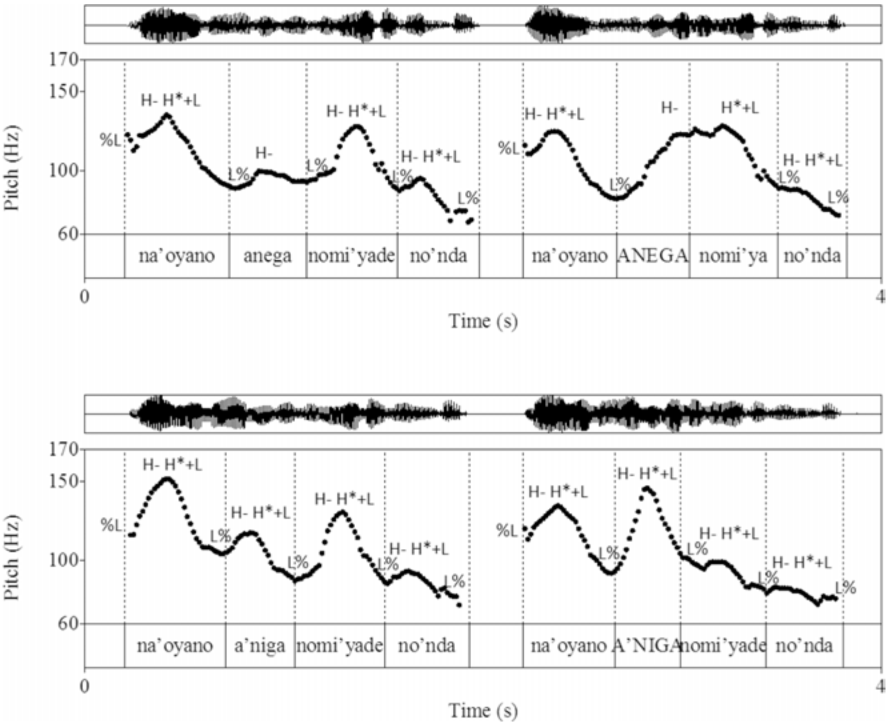


Figure 10: Pitch reset and post-focal compression. *Na'oya-no ane-ga nomi'ya-de no'n-da* Naoya-GEN sister-NOM bar-LOC drink-PAST 'Naoya's sister drank in the bar.', without (top, left) and with (top, right) focus on the second unaccented AP *ane-ga*, and *Na'oya-no a'ni-ga nomi'ya-de no'n-da* Naoya-GEN brother-NOM bar-LOC drink-PAST 'Naoya's brother drank in the bar.', without (bottom, left) and with (bottom, right) focus on the second accented AP *aniga*. Focused words are capitalized.

section 2.7, the pitch ranges of post-focal accented APs are significantly reduced. This process is sometimes called post-focal compression (see also section 4.5). Pitch reset and post-focal compression are shown in Figure 10. The prosodic phrasing in the utterances in this figure is shown in (10), where "{ }" represents the boundaries of the IPs.

(10) The prosodic phrasing in the utterances in Figure 10. Focused words are capitalized.
  a. { (na'oyano) (anega) } { (nomi'yade) (no'nda) } (top, left)
  b. { (na'oyano) } { (ANEGA nomi'yade) (no'nda) } (top, right)
  c. { (na'oyano) (a'niga) } { (nomi'yade) (no'nda) } (bottom, left)
  d. { (na'oyano) } { (A'NIGA) (nomi'yade) (no'nda) } (bottom, right)

The definition of downstep in Japanese differs among researchers. Selkirk and Tateishi (1991) determine the effect of downstep syntagmatically in relation to the

preceding AP. When the F0 peak is lower than that of the preceding AP, they then consider that downstep occurs. When the F0 peak is higher than the preceding one, then downstep is deemed to be blocked and hence pitch range is considered to be reset. Kubozono (1993) on the other hand defines downstep paradigmatically as the lowering effect that is observed in the AP in a sentence that has the same syntactic structure but differs in the accentedness of the preceding AP, regardless of whether the F0 peak of that AP is higher or lower than that of the preceding AP. Both syntagmatic and paradigmatic approaches pose problems in the treatment of downstep and the definition of IP, which is discussed in detail by Ishihara (this volume).

3 Boundary pitch movement

3.1 What is BPM?

As mentioned in the introduction, boundary pitch movements (BPMs) are tones that contribute to the pragmatic interpretation of the utterance. They include a slightly concave rising pitch movement that typically occurs at the end of a question sentence. This type of BPM is transcribed as LH% in X-JToBI. Consider the examples in Figure 11. The LH% BPM can make a sentence ending with a verb in predicative form be interpreted as a question (left), whereas the same sentence is interpreted as a statement without the BPM (right) (Uemura 1989). (Japanese utterances can have no BPM at all.)

It does not follow that LH% always indicates a question. This is shown in Figure 12, where a sentence ending with a verb followed by the sentence-final particle *-yo* remains to be interpreted as a statement even with LH%.

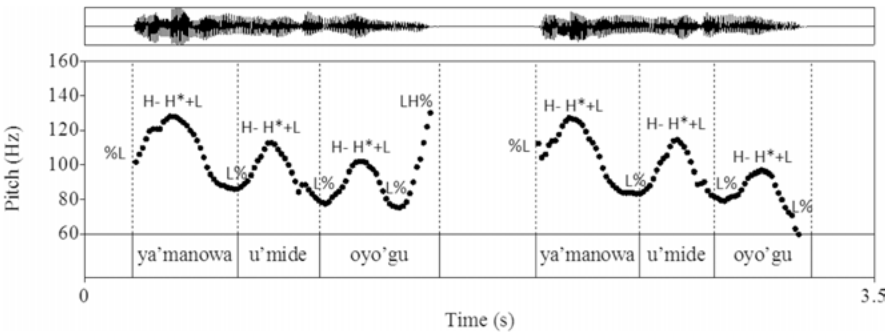


Figure 11: Question utterance with BPM and statement utterance without BPM. *Ya'mano-wa u'mi-de oyo'gu*. (LH%) Yamano-TOP sea-LOC swim 'Will Yamano swim in the sea?' (left) and *Ya'mano-wa u'mi-de oyo'gu*. Yamano-TOP sea-LOC swim 'Yamano will swim in the sea.' (right). The mora assigned a BPM is underlined.

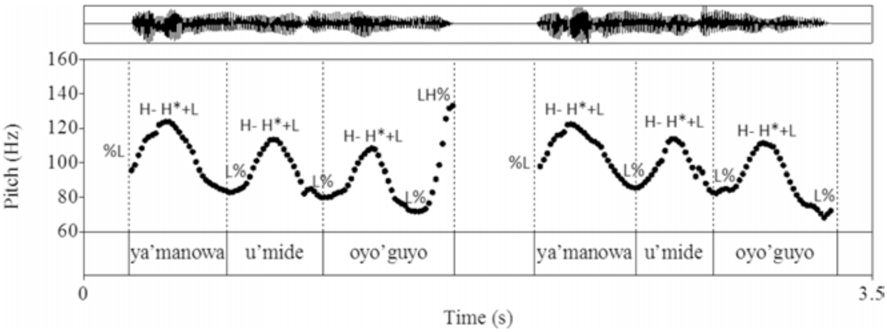


Figure 12: Statement utterances with and without BPM. *Ya'mano-wa u'mi-de oyo'gu-yo*. (LH%) Yamano-TOP sea-LOC swim-SFP 'Yamano will swim in the sea.' (left) and *Ya'mano-wa u'mi-de oyo'gu-yo*. Yamano-TOP sea-LOC swim-SFP 'Yamano will swim in the sea.' (right). The mora assigned a BPM is underlined.

3.2 Inventory of BPMs in X-JToBI

LH% is not the only BPM in the inventory of Japanese BPM. As shown in (11), the inventory of BPMs indicated in the X-JToBI system is H% (simple rise), LH% (scooped rise), HL% (rise-fall) and HLH% (rise-fall-rise), as well as their variations (Maekawa et al. 2002). Each BPM is preceded by the AP-final boundary tone L%.

(11) Four main types of BPMs
   a. H% (Simple rise)
   b. LH% (Scooped rise)
   c. HL% (Rise-fall)
   d. HLH% (Rise-fall-rise)

Figure 13 depicts these four main types of BPM attached to unaccented APs (top) and to accented APs (bottom). As can be seen from the figure, all types of BPM consist of a rise at their beginning. The AP-final L% boundary tone functions as the beginning of the rise, which is in most cases aligned with the onset of the AP-final mora. As is clear from Figure 13, the AP-final L% is not always realized as a low F0, especially in the case of the short unaccented APs (top).

Pierrehumbert and Beckman (1988) assume that BPMs occur only in the sentence-final position and posit the utterance (a prosodic constituent above the IP in their model) as the domain of BPMs. However, BPMs can also occur sentence-medially (Kawakami 1963; Yoshizawa 1960; Uemura 1989; Kori 1997; for the analysis of spontaneous Japanese speech, see Venditti et al. 2008).

Given that BPMs also occur utterance-medially, what is the domain for BPMs? The IP, which is defined as the domain for downstep, should not be the domain for

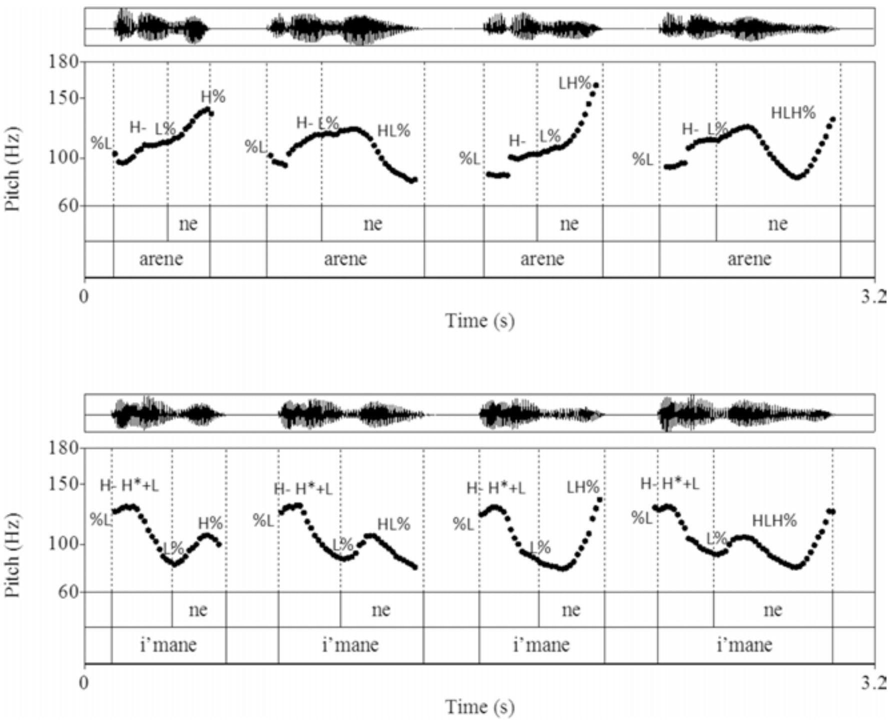


Figure 13: Main types of BPM. Unaccented APs with BPM *Are-ne* that-SFP 'That is...' (top) and accented APs with BPM *I'ma-ne* Now-SFP 'Now...'(bottom). The boundaries of the AP-final mora /ne/ are marked.

BPMs because analysis of the Corpus of Spontaneous Japanese (CSJ, Maekawa 2003), a large scale speech database that contains relatively formal spontaneous speech, revealed that downstep may not be blocked (and hence no pitch reset occurs) even after BPM. Instead, downstep can continue across the phrase boundary to which the BPM is attached (Maekawa et al. 2002). This suggests that the domain for BPMs is a smaller phrasal unit than the IP, and it is the AP within the X-JToBI framework. A different conclusion must be made, however, if we abandon the Strict Layer Hypothesis and adopt the recursive model (Selkirk 2000, 2009; Ito and Mester 2012), in which downstep effects can be nested across phrase to phrase. For the recursive model, see Ishihara (this volume).

Now we move back to the inventory of BPMs. H% differs from LH% mainly in its F0 shape. In the case of H%, F0 starts rising at the beginning of the AP-final mora, whereas in the case of LH% it starts in the middle of the final mora. In addition to this alignment difference, pitch range is generally (but not necessarily) smaller in H% than LH% (Venditti, Maeda, and van Santen 1998; Venditti, Maekawa, and Beckman 2008). The resultant F0 shape for H% is a linear rise with a smaller excursion, while that for LH% is a concave or scooped rise with a larger excursion.

H% in the sentence-final position generally does not cue a question interpretation. Instead, it gives information, for example, that the speaker is insisting (Venditti, Maeda, and van Santen 1998), or that he is firmly persuading the listener to agree with what was said (Uemura 1989). As will be discussed below in section 3.5, H% is sometimes called "emphatic" rise in other frameworks (Kori 1997; Uemura 1989), in that it gives a prominence to the phrase to which H% is attached. H% is also used, according to Uemura (1989), when the speaker is seeking approval, bending the listener to his will, inviting the listener's attention, or blaming. H% can appear sentence-medially, and in this case, H% can also lend prominence to the phrase (Kori 1997; Yoshizawa 1960). It also signals continuation of speech (Kori 1997).

LH% is most often observed at the ends of utterances and typically expresses a question. However, as mentioned in section 3.1, LH% does not always convey a question meaning. Uemura (1989) summarizes the functions of this BPM as an expression of intimacy or a friendly attitude toward the listener. The fact that LH% typically (but not always) occurs at the ends of utterances leads us to speculate that the domain of LH% is higher than that of H% and HL% in the prosodic hierarchy, with the latter two appearing either sentence-medially or sentence-finally. This issue requires further examination in the future.

HL% is a rise-fall BPM, in which the beginning of the rise is at the onset of the AP-final mora with the peak at the end of the rise aligned in the middle of the mora (close to its onset). After the rise, F0 falls at the end of the mora with its duration lengthened. The function of HL% is akin to H%, in that it imparts a prominence to the phrase that the BPM is attached. In their perception study, Venditti, Maeda, and van Santen (1998) revealed that HL% is perceived by the listener as explanatory and emphatic, and it is judged to signal continuation. Citing this study, Venditti, Maekawa, and Beckman (2008) summarize the functions of HL% by saying that listeners expect speakers to use HL% when they are explaining a certain point, and want to focus attention on a particular phrase in this explanation.

The choice between H% and HL% at least partly depends on speaking style and spontaneity. Analysis of the impression rating assigned to the CSJ showed that the rate of H% correlates positively and negatively with speaking style and spontaneity respectively, while the rate of HL% correlates negatively and positively with speaking style and spontaneity (Maekawa 2006). In other words, H% is judged by listeners to be more formal and less spontaneous than HL%.

Not only are their functions similar, but the forms of H% and HL% are also similar to each other. Since the F0 contour of H% is virtually identical to the former part of the contour of HL%, it is possible to hypothesize that H% is a truncated variant of HL%, in which the falling part the rise-fall BPM is curtailed and not realized on the surface because of the short duration of the AP final mora. The fact that the duration of H% is typically shorter than that of HL% increases the plausibility of the hypothesis. Further, H% appearing sentence-medially accompanies an actual

F0 fall after the peak (that is, at the beginning of the next AP), suggesting that H% has a falling property. Since the choice between H% and HL% partly depends on style and spontaneity, it may be reasonable to propose that the (hypothetical) truncated and non-truncated variants (that is, H% and HL%) are in fact stylistic variations of the same type of BPM. This is reminiscent of cross-dialectal variations of British English pitch accents, where the same type of pitch accents are truncated in some dialects but not in the others.[7]

The truncation hypothesis, however, faces difficulties in dealing with the "extended H%" that will be discussed in section 3.4 below. This variant involves the lengthening of the mora that the BPM is attached but accompanies no actual F0 fall after the rise. The absence of the fall in the extended H% regardless of the lengthened duration is not properly explained by the truncation hypothesis, which predicts no truncation when the AP-final mora is lengthened.

The F0 configuration of the former part of HLH% is akin to HL%, but in the case of HLH%, F0 rises again after the fall. The final mora is considerably lengthened. Venditti, Maekawa, and Beckman (2008) suggest that HLH% may be characteristic particularly of infant-directed speech (IDS), where it can give a wheeling or cajoling quality to the utterance to which it is attached. Indeed, the analysis of Japanese infant-directed speech using the RIKEN Japanese Mother-Infant Conversation Corpus (Mazuka, Igarashi, and Nishikawa 2006), which contains IDS and adult-directed speech (ADS) of Japanese, revealed a higher occurrence of HLH% in IDS than ADS (Igarashi et al. 2013). However, this type of BPM occurs much less frequently than other types, even in IDS. It occurs only 12 times in a total of eight hours of IDS produced by 21 mothers. (The frequency of each BPM relative to the total number of APs was, on average per mother, approximately 19.14% for H%, 5.32% for LH%, 2.07% for HL%, and 0.06% HLH% in IDS.) The low frequency of HLH% is also confirmed by the analysis of the CSJ. It occurs only 14 times in the forty-five-hour core portion of the CSJ (Venditti, Maekawa, and Beckman 2008).

3.3 What meanings do BPMs convey?

The preceding subsection provided a brief description of the meanings of BPMs, although it is merely a first approximation. Unfortunately, there is no analysis of the meanings that BPMs convey that is without controversy, and a comprehensive description is beyond the scope of this chapter. However, it is reasonable to point out here that the meanings of Japanese BPMs by and large fit comfortably into Gussenhoven's theory of *biological code*, the theory concerning form-function relations based on the effects of the production process's physiological properties on the speech signal (Gussenhoven 2004).

7 Carlos Gussenhoven suggested this point.

Gussenhoven (2004) identifies three inherent features of the speech production mechanism that affect the rates of vocal fold vibration, which in turn cause variations in pitch. First is larynx size; a smaller larynx produces a higher pitch. Second is the effort expended on speech production; greater effort leads to greater articulatory precision, less undershooting of targets, and hence greater pitch excursion. Third is breathing or the exhalation process; air pressure driving vibratory action becomes smaller along the exhalation phase, yielding high pitch at the beginning and low pitch at the end. Gussenhoven (2004) argues that speakers manipulate the speech production process for communicative purposes, exploiting the correlation between rates of vocal fold vibration and the three biologically determined conditions. The exploitation of the biologically determined effects on pitch variation is called biological code, which is further classified into *Frequency Code*, *Effort Code* and *Production Phase Code*.

Frequency Code, first proposed by Ohala (1984), associates higher or lower pitch with power relations. Effort Code associates wider excursions with greater effort. Production Phase Code associates high pitch with the beginning of utterances and low pitch with the end. Gussenhoven (2004) further claims that the three biological codes explain what is universal about the interpretation of pitch variation. According to Gussenhoven, the general, non-arbitrary form-meaning relation acquires a number of more specific interpretations, which are further classified into 'informational', in which case they signal the attributes of the message, and 'affective', in which case they signal the attributes of the speaker.

Informational interpretations derived from the Frequency Code are, for example, "uncertainty" (for higher pitch) vs. "certainty" (for lower pitch) and hence "questioning" vs. "assertive". Those from the Effort Code are, for example, "more urgent" (for wider excursion) vs. "less urgent" (for smaller excursion), or "more significant" vs. "less significant". Higher and Lower pitch from the Production Code is informationally interpreted as, for example, "new topic" vs. "continued topic" at the beginning of the utterance and as "continuation" vs. "finality" at the end. The main function of LH% is questioning, and it appears to be derived from the Frequency Code. In the same way, some of the meanings of HL% and in particular H% seem to be associated with informational interpretations of the Effort Code, since they are exploited to lead a prominence or emphasis to the constituent of the utterance. Finally, continuation signaled by H% and HL% may be due to informational interpretation of the Production Phase Code.

Affective interpretations deriving from the Frequency Code are, for example, "submissive" vs. "authoritative", "vulnerable" vs. "protective", or "friendly" vs. "not friendly". Those deriving from the Effort Code are, for example, "less surprised" vs. "more surprised", "insistent" vs. 'lacking in commitment' or 'enthusiastic' vs. 'uninterested'. The Production Phase Code is assumed to have informational meanings only. LH% is used to express an intimate or friendly attitude and this function

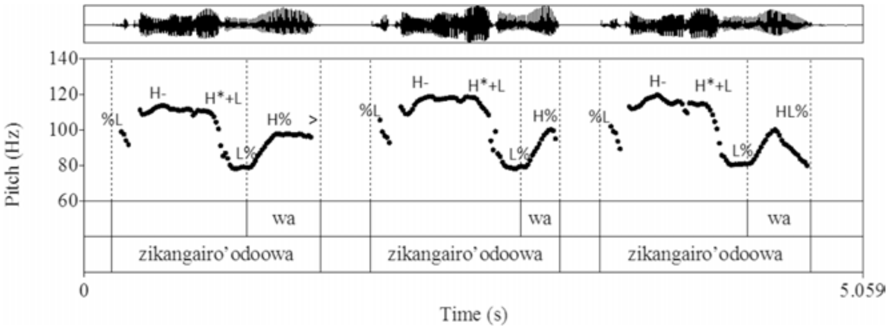


Figure 14: Extended H%> (left), normal H% (middle), and HL% (right). *Zikangairoodoo-wa* overtime. work-TOP. The boundaries of the AP-final mora are marked.

may be derived from the Frequency Code. Meanings that H% conveys such as insistence fit into affective interpretations of the Production Code.

The form-function relation in the biological code is non-arbitrary, and the non-arbitrary form-function relation is generally what we see in regards to intonation, including Japanese intonation.

3.4 Variants of BPMs in X-JToBI

X-JToBI describes types of BPM other than the four types discussed above (H%, LH%, HL%, and HLH%). Here, they are operationally considered to be variants of the main types of BPMs. Future research could reveal that some of them are categorically distinct types.

One of these variants is what X-JToBI regards as a variant of H% and what we might call the "extended H%". In the extended H%, F0 starts to rise at the beginning of the AP-final mora, reaching the peak in the middle of the mora, and then a flat high F0 prevails until the end of the mora. The duration of the mora is lengthened. X-JToBI distinguishes this variant from a simple H% by adding the diacritic ">" or an "extender", and it is tagged as "H%>". An example for the extended H% is illustrated in Figure 14.

The extended H% BPM contrasts with the simple H% by its longer duration and the flat high F0 observed after the rise. It also resembles HL% in its pitch rise around the beginning of the AP-final mora and its lengthened mora duration, but the former differs from the latter in the absence of a subsequent F0 fall.

The other variants involve the dislocation of BPMs; that is, they are realized not in the AP-final mora but in the penultimate. Kawakami (1963) discusses two such types of BPMs; namely what he calls the "hooked rise" (*tsuriagechō*) and the "floating rise" (*ukiagarichō*).

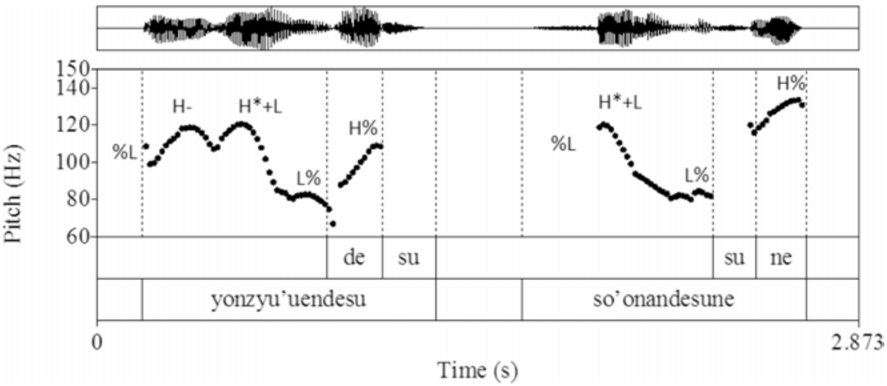


Figure 15: Hooked rise and floating rise. An utterance with the hooked rise *Yonzyu'uen-desu*. forty. yen-COP.POL 'It is forty yen' (left), and an utterance with the floating rise *So'o na-n-desu-ne* so COP-NLZR-COP.POL-SFP 'It is so.' (right). The boundaries of the penultimate and final moras are marked.

The hooked rise is not only defined by its F0 shape, but also by its limited distribution. It only occurs in the sentence-final phrases ending with polite auxiliary verbs *de'su* and *ma'su*. Moreover, devoicing of the final mora /su/ is obligatory for the hooked rise, and thus the rise occurs not in this mora but in the penultimate one. Since the beginning of the rise occurs "around the offset of /de/ or /ma/" (Kawakami 1963), this late alignment of the rise with respect to the mora (late vs. early in the mora) could, along with the dislocation of the rise (penultimate vs. final mora of the AP), be a defining property of the hooked rise that differentiates it from normal H%. Examples of the hooked rise are depicted in Figure 15. X-JToBI considers the hooked rise as a dislocated variant of H%, with the beginning of the rise (L%) aligned around the onset of the penultimate mora and with the end of the rise (H %) aligned at the end of that mora.

It is controversial as to whether the hooked rise is a BPM. Since the auxiliary verbs *de'su* and *ma'su* have a lexical pitch accent in the penultimate mora, the rise can also be analyzed as a manifestation of the peak of the accentual fall produced with an expanded pitch range, with the fall being truncated due to the devoicing of the final mora.

In the floating rise, the beginning and end of the rise are aligned respectively with the onset of the penultimate mora and with the offset of the penultimate mora. Just as in the case of the hooked rise, the floating rise is taken as a variant of H% in X-JToBI. According to Kawakami (1963), the floating rise gives a light quality to the utterance. For example, while the utterance *So'o na-n-desu-ne* 'It is so.' (Figure 15) can be produced with either the floating rise or with (normal) H%, the former signals that the speaker's attitude is lighter (or more imprudent) than the latter. Kawakami (1963) also points out that the floating rise is more likely to occur when

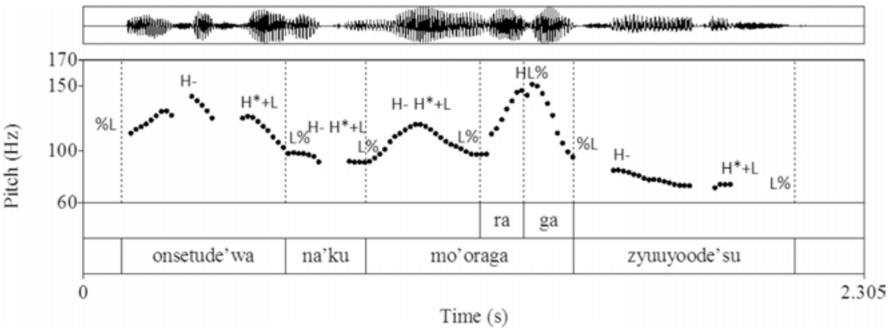


Figure 16: PNLP. *Onsetu-de'-wa na'ku mo'ora-ga zyuuyoo-de'su.* (The mora assigned PNLP is underlined.) 'It is not the syllable but the mora that counts'. The boundaries of the penultimate and final moras (/ra/ and /ga/) of the AP having PNLP are marked.

the last two moras of the AP together constitute a single heavy syllable. When the final heavy syllable is lexically accented, the accent is deleted. For example, the utterance *Bi'iru kudasa'i* 'Give me beer, please!' can be produced with the hooked rise, with the lexical accent in *kudasa'i* deleted.

The difference between the hooked rise and the floating rise discussed below are also controversial in some cases. Kawakami's examples of the floating rise include the utterances ending with the auxiliary verb *ma'su*, whose final mora is devoiced, produced with a rising pitch in the penultimate mora (the accent in *ma'su* is deleted), e.g., *Kore haisyaku deki-masu?* 'Could I borrow this?'. The alignment of the rise may differ between the two types of rises, with it being later for the hooked rise than for the floating rise.

The last dislocated BPM variant is discussed by Oishi (1959), and we refer to this as the penult non-lexical prominence (PNLP) following the terminology in X-JToBI. In PNLP, the F0 rises at the beginning of the AP-penultimate mora with the peak aligned at the end of that mora, and then F0 falls to the end of the AP-final mora so that the AP-penultimate becomes prominent. Figure 16 illustrates PNLP.

On the basis of the analysis of CSJ, Maekawa (2011) revealed that PNLP typically occurs only once in an utterance bounded by strong clause boundaries, and that it occurred most frequently in the penultimate AP of an utterance, suggesting that PNLP is used to predict the end of an utterance.

The lexical unaccentedness of the AP-penultimate mora is a defining property of PNLP in X-JToBI, although formally, the prominence that PNLP creates is almost indistinguishable from that of AP-penultimate pitch accents. They are also functionally similar to each other. In fact, Oishi (1959) does not distinguish PNLP from the prominence found in the AP-penultimate accented mora, such as *otooto'-wa* 'younger brother-top', *mu'gi-sa'e* 'even barley', *kie-re'ba* 'if it disappears…', and *rekisi-to'-wa* 'history-CIT-TOP' (the prominent mora is underlined), while X-JToBI

regards these prominences as a result of a pitch range expansion of the accented AP, distinguishing them from PNLP.

The existence of one or more functional words, such as particles and verbal suffixes, in the AP-final position arguably serves as the necessary condition for the realization of PNLP. A typical condition for PNLP is the AP ending with a noun followed by a single monomoraic particle, such as *mo'ora-ga* 'mora-NOM'. PNLP also occurs in the AP ending with an unaccented bimoraic particle such as *na'goya-kara* 'from Nagoya'. Moreover, it appears in the AP ending with an adjective followed by suffix, such as *ta'kaku-te* 'high'. Oishi (1959) provides only one example in which a PNLP occurs in the AP ending with a noun (without any function word) *moo issyu'ukan* 'one more week', while such cases may be quite rare. Similarly, Maekawa (2011) points out that PNLP can occur in adverbs such as *to'otoo* 'at last' and *so'rosoro* 'gradually', for which case the word accompanies no functional words. This seems to be an exception which may possibly be related to the repetition of the two moras with the same segmental structure, although etiologically they may not be reduplication of the same morphemes.

3.5 How many BPMs are there in Japanese?

No consensus has emerged as to how many BPMs exist in Japanese. Moreover, there are few quantitative analyses of BPMs, a fact that might contribute to the controversy about the inventory of Japanese BPMs. Below we will discuss how researchers agree or disagree on the inventory of BPMs.

Figure 17 is a schematic representation showing categorical boundaries of the rising BPMs that different researchers distinguish. Kawakami's (1963) floating rise and hooked rise are omitted from the figure because here they are regarded as variants of other BPMs.

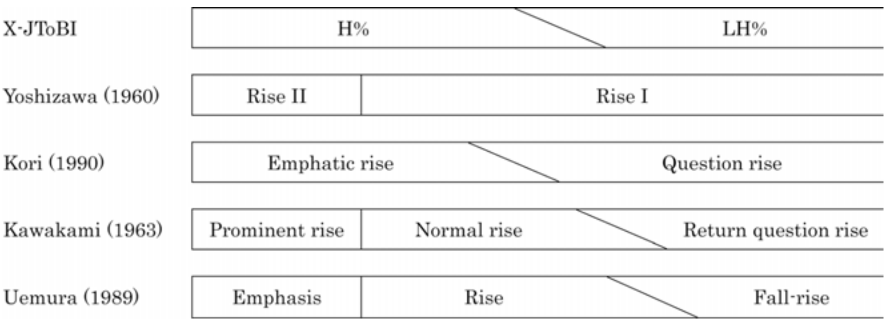


Figure 17: Schematic representation showing correspondences in categories of the rising BPMs identified by different researchers.

Close examination of the previous (mainly qualitative) descriptions of BPMs indicates that most of the researchers distinguish two types of rises. For the sake of convenience they will henceforth be referred to as 'information-seeking question rise' (InfoQ rise) and 'prominence-leading rise' (Prom rise), respectively, following the terminology in Venditti, Maeda, and van Santen (1998). The InfoQ rise is typically observed in a question ending with a verb in predicative form such as *Yameru?* 'Will you quit?'. The Prom rise is typically observed when a speaker is making an insistent statement, such as *Yameru!* 'I will definitely quit!'. The InfoQ rise and Prom rise correspond with, respectively, LH% and H% in X-JToBI.

The InfoQ rise is called "Rise I" (Yoshizawa 1960), "Normal rise" (*futsū no jōshōchō*) (Kawakami 1963), "Question rise" (*gimon jōshōchō*) (Kori 1997), and "Rise" (Uemura 1989). However, Kawakami's "Normal rise" also covers the rising BPM used to give prominence to each phrase, which could be considered as Prom rise here. Thus, the categorical boundary between the two rising BPMs does not always coincide across the compared frameworks. This observation is expressed by slashed lines between the categories in Figure 17. The Prom rise is called "Rise II" (Yoshizawa 1960), "Prominent rise" (*tsuyome no jōshōchō*) (Kawakami 1963), "Emphatic rise" (*kyōchō jōshō*) (Kori 1997), and "Emphasis" (*kyōchō*) (Uemura 1989).

In addition to the two rises discussed above, Kawakami (1963) and Uemura (1989) distinguish the InfoQ rise from what we may call "incredulity question rise" (IncreQ rise), again following the terminology in Venditti, Maeda, and van Santen (1998). IncreQ rise is typically observed in a question where a speaker is expressing disbelief, such as *Yameru??* 'Will you quit??'. Kawakami (1963) terms this type of rise as "Return question rise" (*hanmon no jōhōchō*) and Uemura (1989) as "Fall-rise". X-JToBI does not distinguish between the InfoQ rise and IncreQ rise, and the LH% category covers both.

The IncreQ rise has been described as having a pitch fall before the rise, however to the author's knowledge no experimental results have been reported showing that there is indeed a fall. Instead, the IncreQ rise can best be characterized as having a longer AP-final mora and thus has a longer low-pitched contour than that of the InfoQ rise. Venditti Maeda, and van Santen (1998) conducted an acoustic analysis examining the putative contrast between the InfoQ rise and IncreQ rise and showed that the phonetic distinction between the two question rises was not clear-cut. Their results showed that the final mora on which the rise was realized varies continuously from the shortest (the clearest example of InfoQ rise) to the longest (the clearest example of IncreQ rise). Moreover, the results revealed that the location of the rise onset correlated with the varying durations of the final mora. Venditti, Maekawa, and Beckman (2008) suggest that the InfoQ and IncreQ rises are the extreme endpoints of a continuum that includes many intermediate degrees of emphatic lengthening. Venditti, Maekawa, and Beckman (2008) also point out that the gradient nature of the relationship between the phonetic dimensions and the continuum of contrasting degrees of incredulity suggests an analysis analogous to

the one that Hirschberg and Ward (1995) propose for the uncertainty versus incredulity interpretations of the English fall-rise contour (transcribed as L*+H L- H%).

The inventory of X-JToBI includes no falling BPM, whereas its existence is described in other frameworks (Yoshizawa 1960; Uemura 1989; Kori 1997). In this BPM, F0 decreases sharply in the final syllable accompanied by a lengthening of that syllable. The falling BPM is argued to be used to express, for example, unexpectedness, disgust, contempt, disaffection, and contempt. This BPM occurs only in the utterance that ends with an unaccented or final-accented word (Yoshizawa 1960; Kori 1997). In the other cases, the final syllable is simply lengthened with no F0 fall.

3.6 Variations in contours at the beginning of the utterance

Utterance-initial pitch movements, which are functionally similar to BPMs, are understudied. Kawakami (1956) describes variability in the timing of the initial rise of the utterance-initial AP, showing that the rise aligned earlier or later according to the speaker's emotions. In their experimental studies, Maekawa and Kitagawa showed, among many other things, that the F0 contour at the beginning of the utterance varies significantly depending on the speaker's attitude and intentions ("paralinguistic information" in their terms), such as admiration, disappointment and suspicion (Maekawa and Kitagawa 2002; Maekawa 2004). For example, in the utterance produced with suspicion, the beginning of the initial rise is delayed considerably yielding a long stretch of low F0 before the rise. In addition, the contour exhibits a concave shape in the rising movement (Figure 18). In X-JToBI, this delayed rise is taken as a timing variant of an initial boundary tone (%L), and is tagged by means of a diacritic ">" at the beginning of the rise and %L at the onset of the low F0 region.

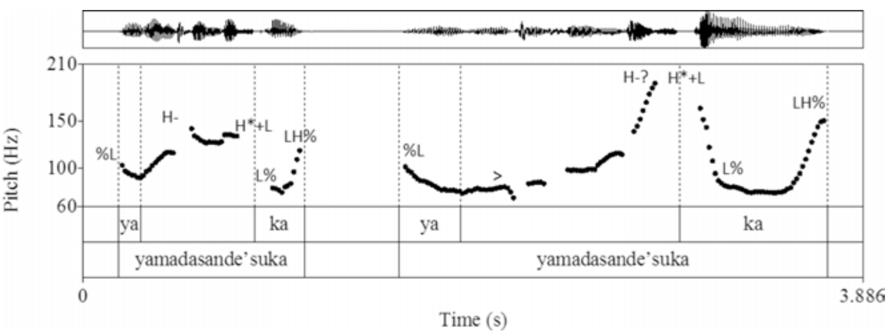


Figure 18: Utterances produced with neutral attitude (left) and with suspicion (right). *Yamada-san-de'su-ka?* Yamada-Mr.-COP.POL-Q 'Is it Mr. Yamada?'. The boundaries of the first and final moras (/ya/ and /ka/) are marked to show their lengthening in the utterance with suspicion.

4 Theoretical implications from a cross-linguistic perspective

4.1 Distribution of AP boundaries with respect to word boundaries

As discussed in section 2.2, an AP can contain two or more lexical words, so that the AP boundaries are distributed sparsely with respect to the number of lexical words in the utterance. This is not necessarily true of all the Japanese dialects. Igarashi (2012, 2014) argues that the distribution of the AP boundaries can be a typological parameter whereby the Japanese dialects are dichotomized into 1) those in which an AP can contain two or more words such as the Tokyo, Koriyama, and Fukuoka dialects and 2) those in which an AP cannot contain two or more words such as the Osaka, Kagoshima, and Kobayashi dialects. In the latter group of dialects, one lexical word followed by a particle in principle constitutes a single AP (In some dialects, including Osaka and Tokyo, some particles form an independent AP.) This typological distinction can also be formulated as the difference in the prosody-syntax mapping rules; namely the Osaka, Kagoshima, and Kobayashi dialects have a prosody-syntax mapping rule whereby the AP boundary is inserted to the left edge of every lexical word, whereas the Tokyo, Koriyama, and Fukuoka dialects do not.

Igarashi (2012) further argues that the typology can be applied to other languages that have the level of AP in their prosodic hierarchy. Seoul Korean (Jun 1998), Bizkaian Basque (Hualde 2003) and French (Jun and Fougeron 2000) are classified into the same group that Tokyo Japanese belongs to, because in these languages the AP boundaries are distributed sparsely with respect to the number of lexical words in the utterance.

In a similar vein, Ladd (1996, 2008) suggests a typology that dichotomizes languages on the basis of the distribution of (intonational) pitch accents in the utterances. Note that the term "pitch accent" here refers to phrasal prominence that is determined by the metrical structure of the phrase, not lexical accent (assigned in the lexicon). In the typology, two groups of languages are identified. One group consists of languages in which almost every lexical word receives a pitch accent. They include, for example, Spanish, Brazilian Portuguese (Elordieta et al. 2003), and Egyptian Arabic (Hellmuth 2007). The other group consists of languages in which only some of the lexical words in the utterance receive pitch accents. They include, for example, English, Dutch (Ladd 1996, 2008), and European Portuguese (Frota 2002).

Extending Ladd's typology, Igarashi (2012) suggests there are languages with sparse tonal distribution, for which tones are sparsely distributed with respect to the number of lexical words in the utterance (such as Tokyo, Koriyama, Fukuoka

Table 1: Classification between languages with dense vs. sparse tonal distribution with respect to the number of words in the utterance proposed by Igarashi (2012). For the distinction between "the phrasing-based" and "accenting-based" frameworks, see text.

	Languages with dense tones	Languages with sparse tones
Languages described in the phrasing-based framework	Osaka Japanese, Kagoshima Japanese, Kobayashi Japanese	Tokyo Japanese, Koriyama Japanese, Seoul Korean, French, Northern Bizkaian Basque, French
Languages described in the accenting-based framework	Spanish, Brazilian Portuguese, Egyptian Arabic	English, Dutch, German, European Portuguese, (French)

Japanese, Seoul Korean, Bizkaian Basque, French, English, Dutch, and European Portuguese) and those with dense tonal distribution, for which tones are densely distributed (such as the Osaka, Kagoshima, and Kobayashi dialects of Japanese, Spanish, Brazilian Portuguese, and Egyptian Arabic) (Table 1). While the usefulness/uselessness of tonal density as a typological parameter is discussed in Hyman (2009), the scope of the discussion is limited to word-prosodic systems, and, to the author's knowledge, there has been no serious study that examines whether phrase-level prosodic systems can be classified on the basis of the density of tones with respect to the number of words in the utterance.

In order to dichotomize languages into those with sparse tones and those with dense tones, it is necessary to propose a new framework that can describe those languages that have the prosodic phrasing at the AP-level but no intonational pitch accents (such as Japanese, Korean and French) and those languages that have intonational pitch accents but no APs (such as English, Dutch, and Spanish).

Beckman and Pierrehumbert (1986) successfully describe Japanese and English, whose intonation systems at first glance differ considerably from each other on the basis of the common framework. Nevertheless, they admit that it is impossible to posit the AP-level prosodic phrasing as found in Japanese for English. Moreover, although their framework captures *structural* similarity between Japanese lexical pitch accents and English intonational pitch accents, these two prosodic entities do not necessarily exhibit *functional* similarity. English pitch accents are assigned post-lexically, whereas Japanese pitch accents are provided in the lexicon, and, unlike in English, they are not deleted by such factors as focus (see section 2.7 above). (The functional differences between the pitch accents of the two languages are emphasized by Venditti, Jun, and Beckman 1996.) In fact, as discussed in the next subsection, English pitch accents are functionally more similar to the AP-level prosodic phrasing in Japanese. We may note, moreover, that English pitch accents are also akin to Japanese BPMs due to the presence of pragmatic contrasts. This is discussed in section 4.3 below.

It would be desirable, therefore, to develop a new model that captures the functional similarities of intonational pitch accent as in English and the AP-level phrasing as in Japanese. For this, it would be necessary to integrate the two different currently accepted frameworks in the AM description of intonation. One is the "phrasing-based framework" that has been applied to languages such as Japanese, Korean, and French (Pierrehumbert and Beckman 1988; Jun and Fougeron 2000), and posits the AP-level phrasing but no intonational pitch accents. The other is the "accenting-based framework" that has been applied to languages such as English, Spanish, and Portuguese (Pierrehumbert 1980; Elordieta et al. 2003), and posits intonational pitch accents but no AP-level phrasing. The integration of the two frameworks will be discussed in the following two subsections.

4.2 Dephrasing vs. deaccenting

Jun (2005) suggests a dichotomy between head-prominence and edge-prominence languages, which roughly corresponds to a distinction between languages that have been described in the accenting-based framework, and those that have been described in the phrasing-based framework, respectively. She points out that there are two ways of prominence realization at a post-lexical level; head prominence and edge prominence. In the former, prominence is realized culminatively by marking the head of a prosodic unit, whereas in the latter, it is realized demarcatively by marking the edge of a prosodic unit. The head-prominence languages include English and other Germanic languages, in which a word becomes prominent by assigning a (intonational) pitch accent to the stressed syllable in the word regardless of its position in the phrase. In the edge-prominence languages, including Japanese and Korean, a phrasal tone marks the edge of a prosodic phrasal unit, and the prominent word comes either at the beginning or the end of the prosodic unit.

It is a common observation that the function of pitch accents in languages such as English (Jun's head-prominence languages) is performed by prosodic phrasing in languages such as Japanese and Korean (Jun's edge-prominence languages). The functions here include the marking of focus. Venditti, Jun, and Beckman (1996) compared the prosodic systems of English, Korean, and Tokyo Japanese. They showed that the function performed by accenting and deaccenting some of the words in English is delivered by inserting and deleting AP boundaries in Korean and Japanese. For example, in English, contrastive focus is roughly realized by a pitch accent followed by deaccenting, while in Korean and Japanese, it is realized by inserting an AP boundary before the focused word, and deleting the AP boundaries of post-focused items. Based on this observation, Venditti, Jun, and Beckman (1996) argue that deaccenting in English produces effects similar to those brought about by dephrasing in Japanese and Korean.

On the basis of the similarity revealed by Venditti, Jun, and Beckman (1996), Ladd (1996, 2008) points out that deaccenting and dephrasing are simply different surface symptoms of the same deeper structural effects. If this is the case, a distinction between head-prominence and edge-prominence could be analyzed as merely two different manifestations of the same abstract structure at a deeper level. For a related discussion, see Truckenbrodt (1995, Chapter 5). Also for a review of his theory, see Ishihara (2011).

4.3 Integration of the phrasing-based and accenting-based frameworks

The possibility that the phrasing-based and accenting-based frameworks are compatible with each other is suggested by Hualde (2003). Hualde analyzes the prosodic systems of the Romance languages including French, Italian, Spanish, Catalan, and Portuguese, and notes that at first glance, the intonation system of French differs greatly from that of the other Romance languages. This impression is brought about by the fact that French intonation has been described in the phrasing-based frameworks (Jun and Fougeron 2000), whereas the other romance languages are described in the accenting-based framework. Hualde (2003) discusses in detail how French differs in prosody from other Romance languages in terms of 1) absence vs. presence of a lexically contrastive accent, 2) anchoring of pitch movements and 3) use of pragmatically contrastive pitch accents. He suggests that those differences are not as profound as the differences that the adopted frameworks imply. Indeed, as Hualde also points out, the intonation of French can be and has been analyzed in the accenting-based framework (e.g., Post 2002).

The fact that French can be described by either framework might suggest that languages such as Japanese and Korean can be described within the same framework as languages such as English and Dutch, with some modifications. The integration would help shed light on what Ladd (1996, 2008) refers to as the same deeper structural effects that underlie different surface manifestations such as dephrasing and deaccenting. Key notions that should be considered are addressed in Hualde's (2003) study concerning Romance languages as cited above.

A prototype of Jun's edge-prominence languages (languages that are typically described within the phrasing-based framework) may be Japanese, in that they exhibit two major characteristics that sharply contrast with the prototype of Jun's head-prominence languages (languages that are typically described within the accenting based framework); namely, locations of post-lexical, or intonational tones and pragmatic contrasts of these tones. Firstly, in Japanese, the locations of intonational tones are restricted to the boundaries of APs. %L and H- are aligned around the phrasal onset, and L% is at the phrasal end. BPMs are linked with the phrase-final mora. Secondly, except for BPMs, intonational tones do not have a pragmatic

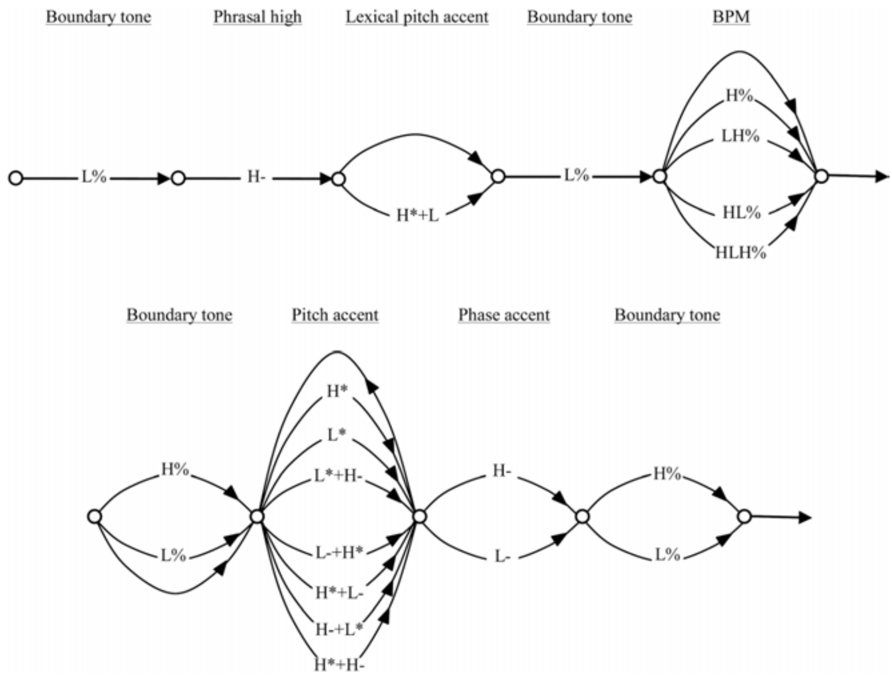


Figure 19: Finite state grammar of intonational contours in Tokyo Japanese (top) adopted from Igarashi et al. (2013) and English (bottom) adopted from Pierrehumbert and Beckman (1988).

choice in their types; they are always L% H- and L%. In other words, pragmatic contrasts in intonational tones are restricted in this language.

In contrast, in English (a prototype of the head-prominence languages), intonational tones appear not only with the phrasal boundaries, but also phrase-medially (Pierrehumbert and Beckman 1988). English boundary tones are linked to the phrasal boundaries, while pitch accents are with the stressed syllable of the words in the phrase. Phrase accent spans the space between the phrase-final pitch accent and the boundary tone. Moreover, all these tones have a pragmatic choice in types; H% vs. L% for boundary tones, H*, L*, L*+H-, L-+H*, H*+L-, H-+L* vs. H*+H- for pitch accents, and H- vs. L- for phrasal accents (Pierrehumbert 1980). With respect to pragmatic contrasts, therefore, English pitch accents are more similar to Japanese BPMs than to Japanese pitch accents.

The fact that pragmatic contrasts in intonational tones are much more limited in Japanese than in English becomes clear by comparing the intonation systems of the two languages described in the form of the finite state grammar in Figure 19. It can be seen that in Japanese, except BPMs, the only tonal options available involve the presence or absence of H*+L. However, this is an intrinsic part of the lexical representation of a word and does not vary according to the speaker's pragmatic intent.

The examination of other languages reveals that there are actually many intermediate types between these prototypical languages in the distribution of intonational tones and their pragmatic contrasts. In Seoul Korean (Jun 2005), one of the languages described within a phrasing-based framework, tones are aligned with phrasal edges as in Japanese. Specifically, the tones marking the AP are THLH, where T indicates either the H or L tone, which is determined by the laryngeal feature of the AP-initial consonant. The first two tones are aligned with the phrasal onset, while the last two with the phrasal end. The pragmatically contrastive tones are, again just as in Japanese, limited to BPMs (termed "boundary tones" in Jun's framework). The Korean system, therefore, appears to be almost identical to that of Japanese with respect to both restricted distribution of intonational tones (only at the phrasal edges) and their restricted pragmatic contrasts (only for BPMs). However, the middle two tones of the AP-marking THLH may actually be unrealized, resulting, for example, in a LHH or LLH tone sequence. The choice of tones varies across speakers and across different discourse contexts, while as Jun (2005) pointed out, it is not exactly clear what the conditions for these variations are. Further research may be needed to examine whether or not they involve pragmatic contrasts, though Jun (2005) states that different realizations do not seem to have contrastive meaning. If they do, then the additional contrast in tones makes Seoul Korean closer to English, a language that has a number of pragmatically contrastive tones.

Similarly, French, another language described within the phrasing-based framework (Jun and Fougeron 2000), is analyzed to have the AP-marking tones LHiLH*. The distribution of these tones is restricted just as in Japanese and Korean, in that they are all aligned with the phrasal edges. The high tone with a star (H*) represents the tone associated with the phrase-final full vowel. The high tone with lower case 'i' (Hi) indicates a tone that is associated with the first syllable of the AP-initial lexical word but which can be realized on the preceding or following syllable. Pragmatic choices in tones are, just as in Japanese and Korean, limited to BPMs (termed "boundary tones" in this framework). Like Seoul Korean, however, the AP-marking tones do exhibit variations, being realized as, for example, LLH*. Jun and Fougeron (2000) do not see pragmatic relevance in these variations. As mentioned in section 4.3 above, however, French is also described within the accenting-based framework (e.g., Post 2002), in which the variation in tones of the AP is considered to have different pragmatic meanings. Thus, with respect to the number of pragmatic tone contrasts, French might be considered to be located at the midpoint in the continuum between a prototypical edge-prominence language and a prototypical head-prominence language.

In Egyptian Arabic, which has been analyzed using the accenting-based framework, the distribution of intonational tones is not limited to prosodic phrase boundaries (Hellmuth 2007). In this respect, Egyptian Arabic is similar to English. However, this language, unlike languages such as English, has no choice in pitch accents; it is

always LH*. Pragmatic contrast is limited to tones that mark prosodic phrase boundaries; L- vs. H- for phrasal tones, and L% vs. H% for boundary tones. Egyptian Arabic thus shares a property with Japanese, that is, the limitation in pragmatic contrasts in intonation tones.[8]

In summary, the comparison of the prosodic structure in several languages suggests that distinction between edge-prominence and head-prominence languages is not categorical; there appear to be a number of intermediate types between the two extremes in the continuum. This in turn suggests that the phrasing-based and the accenting-based frameworks in the description of intonation systems might not reflect substantial cross-linguistic variation.

4.4 Cross-linguistic difference in post-focal compression

Languages may differ in the prosodic realization of focus. In addition to the difference between dephrasing and deaccenting discussed above, some languages are reported to lack prosodic means of signalling focus. Below we will see that the distinction that Igarashi (2012) proposed between languages with sparse tones and those with dense tones (discussed in section 4.1) will help us to capture cross-linguistic differences in the relationship between prosodic and focus structures.

In languages with sparse tones, such as English, Dutch, French, Seoul Korean, and Tokyo Japanese, differences in the focus structure can be signaled by means of deaccenting or dephrasing (Ladd 1996, 2008; Jun and Fougeron 2000; Jun 2005; Venditti, Jun, and Beckman 1996; Pierrehumbert and Beckman 1988). In languages with dense tones, the differences are not signaled by deaccenting or dephrasing, since every lexical word generally receives a pitch accent or is phrased into an AP. A primary prosodic device that signals focus in these languages would therefore instead be pitch range modification of post-focal words.

Pitch range reduction of post-focal words is known as a post-focal compression (PFC), and PFC is a primary prosodic means of signalling focal structure in the languages with dense tones (Igarashi 2012). Crucially, PFC is virtually the only prosodic manifestation of focus in arguably most of the languages with dense tones including the Osaka dialect of Japanese (Kori 1987; Pierrehumbert and Beckman 1988; Igarashi 2014).

It must be pointed out here that PFC is not a property limited to languages with dense tones. The languages with sparse tones appear to exhibit both means, that is, dephrasing/deaccenting as well as PFC. English and Tokyo Japanese, for example,

8 In addition, as discussed in section 4.1, every content word receives a pitch accent in Egyptian Arabic. Together with the absence of contrasts in pitch accents, this leads to the question of whether pitch accents in this language should be taken as intonational tones. Rather, they would be best treated as word-level tones.

both exhibit PFC (Pierrehumbert 1980; Ladd 1996, 2008; Pierrehumbert and Beckman 1988; Venditti, Jun, and Beckman 1996; Sugahara 2003).

A typologically intriguing fact is that PFC is not universal. Xu (2011) demonstrates that Chinese dialects (typical dialects/ languages with dense tones) can differ in whether or not they exhibit PFC. Mandarin has PFC whereas Taiwanese and Cantonese do not. Several languages other than Chinese lack PFC, including Yucatec Maya, Chichewa, and Hausa (see references in Xu 2011). Xu, Chen, and Wang (2012) point out that the absence of post-focal compression in a given language is not related to the presence of lexically contrastive tones. Mandarin exhibits PFC but Taiwanese does not, although both have lexically contrastive tones. Further, Wang, Wang, and Qadin (2011) show that Wa and Deang (Mon-Khmer languages) have no lexical tones but lack post-focal compression. Xu, Chen, and Wang's (2012) claim is strengthened by the observation that languages such as Egyptian Arabic (Hellmuth 2007), Seoul Korean, and Kobayashi Japanese (Igarashi 2014) have no lexically contrastive tones but exhibit PFC.

Given that the presence/absence of PFC is independent from the presence/ absence of lexically contrastive tones, it is interesting to raise the question as to whether the presence/absence of PFC correlates with the distinction between languages with sparse vs. dense tones. Igarashi (2012) explores this possible correlation and points out that there appears to be no language that has sparse tonal distribution but no PFC. Based on this observation he suggests that there might be an implicational relationship between the density of tones and the presence/absence of PFC; namely, if a language has no PFC, then the language has dense tones, but not vice versa.

5 Conclusion

This chapter has provided an overview of the intonation system of Japanese. Since the discussion concerning prosodic phrasing was based on the X-JToBI framework, it has been impossible to fully cover some recent theories of prosodic phrasing, especially the recursive model (Selkirk 2000, 2009; Ito and Mester 2012), which arguably has an advantage over the frameworks with the prosodic hierarchy that obeys the Strict Layer Hypothesis. Readers may refer to Ishihara (this volume) for discussion on the recursive model from the viewpoint of prosody-syntax interface.

Forms and functions of BPMs discussed in Section 3 clearly need further investigation. Quantitative analysis of BPMs, including perception and production experiments (Venditti 1998), as well as analysis of the speech corpus (Maekawa 2011) will contribute to addressing these understudied issues.

From a cross-linguistic point of view, Japanese provides intriguing ground. Similarities between dephrasing and deaccenting, a typological distinction between

edge-prominence and head-prominence languages, and cross-linguistic differences of post-focal compression discussed in section 4 are merely a few of the points of interest.

Acknowledgements

For their collaboration and feedback, I wish to thank Akira Utsugi, Carlos Gussenhoven, Dona Erickson, Kan Sasaki, Shigeto Kawahara, and Shinichiro Ishihara. The study reported in this paper was supported in part by Grant-in-Aid for Young Scientists B #23720207.

References

Akinaga, Kazue. 2002. *Shinmeikai nihongo akusento jiten* [Shinmeikai Japanese accent dictionary]. Tokyo: Sanseido.

Beckman, Mary, Julia Hirschberg and Stefanie Shattuck-Hufnagel. 2006. The original ToBI system and the evolution of the ToBI Framework. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 7–54. Oxford: Oxford University Press.

Beckman, Mary and Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bolinger, Dwight. 1972. Accent is predictable (if you're a mind-reader). *Language* 48. 633–644.

Bruce, Gøsta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Elordieta, Gorka, Sónia Frota, Pilar Prieto and Marina Vigário, M. 2003. Effects of constituent length and syntactic branching on intonational phrasing in Ibero-Romance. *Proceedings of the 15th International Congress of Phonetic Sciences*, 487–490.

Frota, Sónia. 2002. Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation. *Probus* 14. 113–146.

Fujisaki, Hiroya. 1989. Nihongo no onchō no bunseki to moderuka: Goakusento, tōgo kōzō, danwa kōzō to onchō to no kankei [Modeling of prosody and syntax in Japanese: Relation between word accent, syntactic structure, and prosody]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 266–297. Tokyo: Meiji Shoin.

Fujisaki, Hiroya and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan 5*(4). 233–242.

Fujisaki, Hiroya and Hiroshi Sudo. 1971. Synthesis by rule of prosodic features of connected Japanese. *Proceedings of the 7th International Congress on Acoustics*, Budapest, 133–136.

Grabe, Esther, Brechtje Post, Francis Nolan and Kimberley Farrar. 2000. Pitch accent realization in four varieties in British English. *Journal of Phonetics* 28. 161–185.

Grice, Martine. 1995. *The intonation of interrogation in Palermo Italian: Implications for intonation theory*. Tubingen: Niemeyer.

Grice, Martine, Stefan Baumann and Ralf Benzmüller. 2006. German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford: Oxford University Press.

Gussenhoven, Carlos 1991. The English rhythm rule as an accent deletion rule. *Phonology* 8(1). 1–35.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Hattori, Shiro. 1954. On'inron kara mita kokugo no akusento [The accent system of the national language from a phonological point of view]. *Kokugogaku Kenkyū* [Studies in Japanese linguistics] 2. Reprinted in Shiro Hattori 1960. *Gengogaku no hōhō* [Methodologies in linguistics], 240–277. Tokyo: Iwanami.

Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.

Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review* 24(2–3). 291–316.

Hirschberg, Julia and Gregory Ward. 1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24(4). 407–412.

Hualde, José Ignacio. 2003. Remarks on the diachronic reconstruction of intonational patterns in Romance with special attention to Occitan as a bridge language. *Catalan Journal of Linguistics* 2. 181–205.

Hyman, Larry M. 2009. How (not) to do phonological typology. *Language Science* 31. 213–238.

Igarashi, Yosuke. 2012. Typology of prosodic phrasing in Japanese dialects from a cross-linguistic perspective. *Lingua* 122. 1441–1453.

Igarashi, Yosuke. 2014. Typology of intonational phrasing in Japanese dialects. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 464–492. New York: Oxford University Press.

Igarashi, Yosuke, Hideaki Kikuchi and Kikuo Maekawa. 2006. Inritsu jōhō [Prosodic information]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kokuritsu Kokugo Kenkyūjo hōkoku 124: Nihongo hanashi kotoba kōpasu no kōchikuhō* [Construction of the Corpus of Spontaneous Japanese], 347–453. Tokyo: National Institute for Japanese Language.

Igarashi, Yosuke, Ken'ya Nishikawa, Kuniyoshi Tanaka and Reiko Mazuka. 2013. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134(2). 1283–1294.

Ishihara, Shinichiro. 2007. Major Phrase, Focus Intonation, Multiple Spell-Out (MaP, FI, MSO). *The Linguistic Review* 24. 137–167.

Ishihara, Shinichiro. 2011. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121(13). 1870–1889.

Ishihara, Shinichiro. this volume. Chapter 14: Syntax-phonology interface. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ito, Kiwako. 2002. T*he interaction of focus and lexical pitch accent in speech production and dialogue comprehension: Evidence from Japanese and Basque*. Urbana-Champaign, Illinois: University of Illinois dissertation.

Ito, Junko and Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya and Mariko Sugahara (eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk*, 280–303. London: Equinox Publishers.

Ito, Junko and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124(1). 20–40.

Jun, Sun-Ah. 1998. The accentual phrase in Korean prosodic hierarchy. *Phonology* 15. 189–226.

Jun, Sun-Ah. 2005. Prosodic typology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 430–458. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah and Cécile Fougeron. 2000. A phonological model of French intonation. In Antonis Botinis (ed.), *Intonation: Analysis, modelling and technology*, 209–242. Dordrecht: Kluwer.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto and Takahito Shinya. 2008. The intonation of gapping and coordination in Japanese: Evidence for intonational phrase and utterance. *Phonetica* 65. 62–105.

Kawakami, Shin. 1956. *Buntō no intonēshon* [Sentence-initial intonation]. Kokugogaku 25. 21–30.

Kawakami, Shin. 1957a. Tōkyōgo no takuritsu kyōchō no onchō [Tonal prominence in Tokyo Japanese]. Reprinted in Shin Kawakami 1995. *Nihongo akusento ronshū* [Papers on Japanese accent], 76–91. Tokyo: Kyūko Shoin.

Kawakami, Shin. 1957b. Jun'akusento ni tsuite [On quasi-accent]. *Kokugo Kenkyū* [Study of the National Language] 7. 21–31.

Kawakami, Shin. 1963. Bunmatsu nado no jōshōchō ni tsuite [On sentence-final pitch rises]. Reprinted in Shin Kawakami 1995. *Nihongo akusento ronshū* [Papers on Japanese accent], 274–298. Tokyo: Kyūko Shoin.

Kindaichi, Haruhiko. 1951. Kotoba no senritsu [The Rhythm of Language]. *Kokugogaku* [Studies in the Japanese language] 5(2). 37–59.

Kitagawa, Yoshihisa 2005. Prosody, syntax and pragmatics of wh-questions in Japanese. *English Linguistics* 22(2). 302–346.

Kori, Shiro. 1987. The tonal behavior of Osaka Japanese: An interim report. *Ohio State University Working Papers in Linguistics* 36. 31–61.

Kori, Shiro. 1997. Nihongo no intonēshon: Kata to kinō [Japanese intonation: Its patterns and functions]. In Tetsuya Kunihiro, Hajime Hirose and Morio Kohno (eds.), *Nihongo onsei 2: Akusento intonēshon rizumu to pōzu* [Sound in Japanese 2: Accent, intonation, rhythm, and pause], 169–202. Tokyo: Sanseido.

Kubozono, Haruo. 1988. *The organization of Japanese prosody*. Edinburgh: University of Edinburgh dissertation. (Kurosio, 1993).

Kubozono, Haruo. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6. 39–67.

Kubozono, Haruo 2007. Focus and intonation in Japanese: Does focus trigger pitch reset? In Shinichiro Ishihara (ed.), *Interdisciplinary studies on information structure*, 1–27. Potsdam: Universitätsverlag Potsdam.

Ladd, D. Robert. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.

Ladd, D. Robert. 1988. Declination 'reset' and the hierarchical organization of utterance. *Journal of the Acoustical Society of America* 84. 530–544.

Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert. 2008. *Intonational phonology*, 2nd edition. Cambridge: Cambridge University Press.

Maekawa, Kikuo. 1994. Is there 'dephrasing' of the accentual phrase in Japanese? *Ohio State University Working Papers in Linguistics* 44. 146–165.

Maekawa, Kikuo. 1997. Nihongo gimonshi gimonbun no intonēshon [Intonation in Japanese wh-questions]. *Bunpō to Onsei* [Grammar and sound], 45–53. Tokyo: Kurosio.

Maekawa, Kikuo. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition* 4. 7–22.

Maekawa, Kikuo. 2004. Production and perception of 'paralinguistic' information. *Proceedings of Speech Prosody 2004*, Nara, Japan, 367–374.

Maekawa, Kikuo. 2006. Analysis of language variation using a large-scale corpus of spontaneous speech. *Proceedings of the International Symposium on Linguistic Patterns in Spontaneous Speech*, Taipei, 15–37.

Maekawa, Kikuo. 2011. PNLP no onseiteki keijō to gengoteki kinō [Phonetic shape and linguistic function of penultimate non-lexical prominence]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 15(1). 1–13.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti. 2002. X-JToBI: An extended J_ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing* (ICSLP 2002), Denver, Colorado, 1545–1548.

Maekawa, Kikuo and Norimichi Kitagawa. 2002. Onsei wa paragengo jōhō o ikani tsutaeru ka [How Does Speech Transmit Paralinguistic Information?]. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* 9(1). 46–66.

Maekawa, Kikuo and Yosuke Igarashi. 2007. Prosodic phrasing of bimoraic accented particles in spontaneous Japanese. *Proceedings of the 17th International Congress of Phonetic Science*, Saarbrücken, Germany, 1217–1220.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Mazuka, Reiko, Yosuke Igarashi and Kenya Nishikawa. 2006. Input for learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus. *Technical report of IEICE*, TL2006–16, 106(165). 11–15.

Miyata, Koichi. 1928. Nihongo no akusento ni kansuru watashi no kenkai [My opinions about the accent system of Japanese]. *Onsei no Kenkyū* [The Study of Sounds, The Phonetic Society of Japan] 2. 31–37.

Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Ohala, John, J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41. 1–16.

Oishi, Hatsutaro. 1959. Purominensu ni tsuite: Tōkyōgo no kansatsu ni motozuku oboegaki [On prominence: Notes based upon the observation of Tokyo Japanese]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kotoba no kenkyū* 1, 87–102. Tokyo: NLRI.

Okumura, Mitsuo. 1956. Ji no keitaironteki seikaku [Morphological nature of affixes]. *Kokugo Kokubun* 25(9). 539–552.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT dissertation.

Pierrehumbert, Janet and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge: MIT Press.

Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.

Post, Brechtje. 2002. *French tonal structures. Proceedings of the Speech Prosody 2002*, Aix-en-Provence, 583–586.

Sagisaka, Yoshinori and Hirokazu Sato. 1983. Accentuation rules for Japanese word concatenation. *Transactions of the Institute of Electronics and Communication Engineers of Japan* J66-D(7). 849–856.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.

Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In James L. Morgan and Katherine Demuth (eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In Merle Horne (ed.), *Prosody: Theory and experiment*, 231–261. Dordrecht: Kluwer Academic Publishers.

Selkirk, Elisabeth O. 2009. On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Journal of the Linguistics Society of Japan* 136. 1–39.

Selkirk, Elisabeth O. and Koichi Tateishi. 1991. Syntax and downstep in Japanese. In Carol Georgopoulos and Roberta Ishihara (eds.), *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda*, 519 –543. Dordrecht, Boston: Kluwer Academic.

Shattuck-Hufnagel, Stephanie and Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Silverman, Kim, Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert and Julia Hirschberg 1992. ToBI: A standard for labeling English prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, Canada, 867–870.

Sugahara, Mariko. 2003. *Downtrends and post-FOCUS intonation in Tokyo Japanese*. Amherst, MA: University of Massachusetts dissertation.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus, and prominence.* Cambridge, MA: MIT dissertation.

Uemura, Yukio. 1989. Nihongo no intonēshon [Intonation in Japanese]. In Gengogaku Kenkyūkai (ed.), *Kotoba no kagaku vol. 3* [Science of language vol. 3], 193–220. Tokyo: Mugi Shobō.

Uyeno, Tazuko, Hideo Hayashibe and Kunihiko Imai. 1979. On pitch contours of declarative, complex sentences in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 13. 175–187.

Uwano, Zendo. 1989. Nihongo no akusento [Accent in Japanese]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 178–205. Tokyo: Meiji Shoin.

Uwano, Zendo. 1999. Classification of Japanese accent systems. In Shigeki Kaji (ed.), *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics'*. 151–178. Tokyo: ILCAA.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2). 219–255.

Vance, Timothy J. 2008. The sounds of Japanese. New York: Cambridge University Press.

Venditti, Jennifer. 2005 The J_ToBI model of Japanese intonation. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*. 172–200. Oxford: Oxford University Press.

Venditti, Jennifer, Kazuaki Maeda and Jan P. H. van Santen. 1998. Modeling Japanese boundary pitch movements for speech synthesis. *Proceedings of the Third ESCA Workshop on Speech Synthesis*, Jenolan Caves, Australia, 317–322.

Venditti, Jennifer, Kikuo Maekawa and Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 456–512. New York: Oxford University Press.

Venditti, Jennifer, Sun-Ah Jun and Mary E. Beckman. 1996. Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 287–311. Mahwah, NJ: Lawrence Erlbaum.

Yoshizawa, Norio. 1960. Intonēshon [Intonation]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kokuritsu Kokugo Kenkyūjo hōkoku 18: Hanashi kotoba no bunkei 1: Taiwa shiryō ni yoru kenkyū* [Research report 18: Sentence patterns in Japanese 1: Research based on conversational speech], 249–288. Tokyo: Shūei Shuppan.

Wang, Bei, Ling Wang and Tursun Qadir. 2011. Prosodic realization of focus in six languages/dialects in China. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 144–147.

Xu, Yi. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. *Proceedings of the 17th International Congress of Phonetic Science*s, Hong Kong, 152–155.

Xu, Yi, Szu-Wei Chen and Bei Wang. 2012. Prosodic focus with and without Post-Focus Compression (PFC): A typological divide within the same language family? *The Linguistic Review* 29(1). 131–147.

Shinichiro Ishihara

14 Syntax–phonology interface

1 Introduction

This chapter surveys theoretical and empirical issues related to the syntax–phonology interface in Japanese.[1] In this chapter, the term *syntax–phonology interface* will be used to refer to all areas of linguistic research which deal with either phonology or syntax, and with how either of these components of the grammar interacts with the other one. With this definition, studies of the syntax–phonology interface may be divided into two groups, depending on the researcher's perspective. With a *phonological perspective*, one aims to develop a phonological theory that accounts for the effects of syntax on prosody. With a *syntactic perspective*, on the other hand, one aims to account for syntactic phenomena by taking phonological information into account. This chapter mainly discusses studies with a phonological perspective, while only briefly referring to several studies with a syntactic perspective.

Most phonological theories of the syntax–phonology interface assume different types of requirements that a prosodic structure needs to fulfill. The prosodic structure created for an utterance is then a result of the interaction of these requirements. In this chapter, three types of requirements will be discussed: i) syntax–prosody mapping, ii) prosodic wellformedness, and iii) information structure–prosody mapping. First, the set of *syntax–prosody mapping* principles requires that the syntactic information of a sentence be represented in (or mapped onto) the prosodic structure, so that it can be transmitted to the hearer through prosody. Second, prosodic structures are subject to various types of *wellformedness* conditions. In Japanese, for example, lexical pitch accents impose certain conditions on the realization of the prosodic structure. Third, the *information structure* of the sentence imposes further restrictions on prosody. These restrictions sometimes interfere with each other. The goal of the studies of the syntax–phonology interface (with a phonological perspective) is, therefore, to disentangle the influences of syntactic, prosodic, and information-structural factors on prosody, and to understand how they interact with each other.

The chapter is organized as follows. Section 2 establishes the theoretical background of the chapter. The theoretical development of the prosodic hierarchy in Japanese will be reviewed. Section 3 discusses theories of the syntax–prosody mapping. The interaction between the mapping principles and the prosodic wellformedness conditions will also be illustrated. Section 4 discusses the information structure–prosody mapping, i.e., how the prosodic realization of discourse information interacts with the syntax–prosody mapping and prosodic wellformedness.

1 Throughout this chapter, the discussion is based on the intonation system of Tokyo Japanese, unless otherwise noted.

Section 5 surveys some of the syntax–prosody interface studies that have taken a syntactic perspective. Section 6 summarizes the chapter, and presents remaining issues.

2 The prosodic hierarchy of Japanese

Many phonological theories of intonation assume some form of *prosodic structure* that is independent of the syntactic structure (Pierrehumbert 1980; Selkirk 1984, 1986; Nespor and Vogel 1986). Which prosodic categories are assumed in the *prosodic hierarchy* (the hierarchical organization of prosodic categories), as well as the names for these categories, varies notoriously. Researchers argue for different numbers of levels, and sometimes use the same term to refer to different levels.

Table 1 summarizes the terminologies used by major studies. Researchers are divided into three groups: (i) researchers working under the Autosegmental–Metrical framework (Pierrehumbert and Beckman 1988) and/or ToBI frameworks (Venditti 2005; Maekawa et al. 2002; Venditti, Maekawa, and Beckman 2008; Igarashi, this volume), (ii) researchers adopting McCawley's (1968) terminology (McCawley 1968; Poser 1984; Kubozono 1993; Kawahara and Shinya 2008), and (iii) researchers adopting the *Syntax–Prosody Mapping Hypothesis* (*SPMH*, see section 3) (Ito and Mester 2007, 2012, 2013; Selkirk 2009, 2011).

This chapter will adopt the terminology of the third group, more specifically, the one proposed by Ito and Mester (2013), given in (1). (The terms Minor and Major Phrases will also be used, whenever necessary, to facilitate the discussion.) The

Table 1: Terminologies of prosodic constituents in the major literature on Japanese intonation.

<table>
<tr><td colspan="2">AM theory / ToBI</td><td colspan="2">Minor/Major Phrase</td><td>Syntax–Prosody Mapping</td></tr>
<tr><td>Pierrehumbert & Beckman</td><td>J_ToBI, X-JToBI</td><td>McCawley, Poser, Kubozono</td><td>Kawahara & Shinya</td><td>Ito & Mester, Selkirk, this chapter</td></tr>
<tr><td rowspan="2">Utterance</td><td rowspan="3">Intonation Phrase</td><td rowspan="2">(not discussed)</td><td>Utterance</td><td>ι_{max}</td></tr>
<tr><td>Intonational Phrase</td><td>PClause or Intonational Phrase (ι)</td></tr>
<tr><td>Intermediate Phrase</td><td>Major Phrase</td><td>Major Phrase</td><td>Phonological Phrase (φ)</td></tr>
<tr><td>Accentual Phrase</td><td>Accentual Phrase</td><td>Minor Phrase</td><td>Minor Phrase</td><td>φ_{min}</td></tr>
</table>

top two levels, *PClause* and *PPhrase*, are particularly relevant for the study of the syntax–phonology interface in Japanese. In order to motivate this particular set of prosodic categories, it is necessary to survey the history of intonation theories for Japanese.

(1) Prosodic hierarchy in Japanese (Ito and Mester 2013: 26)
 a. Phonological Clause / PClause (ι)[2]
 b. Phonological Phrase / PPhrase (φ)
 c. Phonological Word / PWord (ω)

2.1 PPhrase

In the studies in the early stages of generative grammar (McCawley 1968; Haraguchi 1977; Poser 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988), the exact mechanism of the syntax–prosody mapping had not yet been established. Their major contribution is the analysis of the tonal behavior of lexical pitch accents, and their interaction with other lexical and post-lexical rules, such as compound accent rules and rules governing prosodic phrase formation (see Kawahara, this volume for the phonology of pitch accents).

Most theories of Japanese intonation have assumed two distinct prosodic categories for what this chapter refers to as the PPhrase (cf. Igarashi, this volume). These prosodic categories have been called *Minor Phrase* and *Major Phrase* (also known as *Accentual Phrase* and *Intermediate Phrase*, respectively). McCawley (1968) first introduced these two prosodic domains based on the realization of lexical pitch accents. Three prosodic cues have been relevant to the definition of Minor and Major Phrases: accent culminativity, initial rise, and downstep.

2.1.1 Accent culminativity

McCawley (1968) defined the Minor Phrase as the domain of pitch accent realization. Within a Minor Phrase, *at most one* lexical pitch accent may be realized. In other words, a Minor Phrase may contain maximally one accented word. This restriction is called *accent culminativity* (Ito and Mester 2012, 2013).

Lexical items in Japanese can be divided into accented and unaccented ones. In Japanese, a pitch accent is always realized with a falling F0 contour (labeled "H*+L" in the ToBI transcription system, Venditti 2005; Maekawa et al. 2002; Venditti,

2 This level is more commonly called "Intonational Phrase" (hence "ι" for short). This chapter adopts a rather uncommon term "PClause" from Ito and Mester (2013). See Ito and Mester (2013: fn.6) for a theoretical motivation of this term.

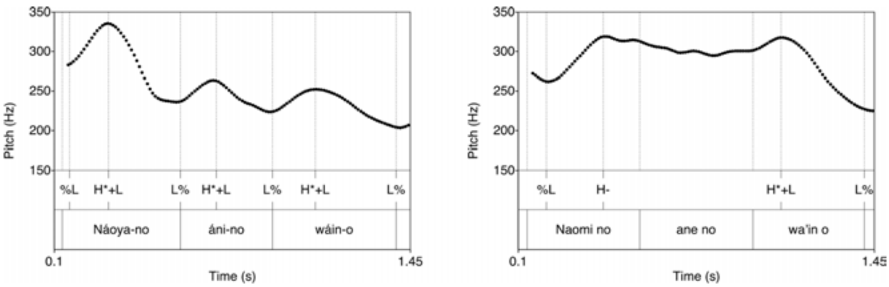


Figure 1: Sample F0 contours of (2a) (left) and (2b) (right).

Maekawa, and Beckman 2008; Igarashi, this volume), which starts on the mora specified as lexically accented. Therefore, accented words always contain the accentual F0 fall, while unaccented words do not contain any such fall.

(2a) is composed of three accented words, whereas (2b) is composed of a sequence of two unaccented words followed by an accented word. (In the examples in this chapter, the vowel bearing a pitch accent is indicated by an apostrophe following it.) An accentual F0 fall can be observed in each word in the former, while only the last word exhibits a fall in the latter, as shown in Figure 1:[3]

(2) a. *Na'oya no a'ni no wa'in o* 'Naoya's big brother's wine ACC'
b. *Naomi no ane no wa'in o* 'Naomi's big sister's wine ACC'

Accent culminativity prohibits two accented words from forming a single Minor Phrase. In (2a), therefore, each accented word ω_A forms its own Minor Phrase, i.e., $(\omega_A)_{MiP}$ $(\omega_A)_{MiP}$ $(\omega_A)_{MiP}$. As in (2b), in contrast, more than one unaccented word ω_U plus at most one accented word may (but do not have to) form a single Minor Phrase together, $(\omega_U\ \omega_U\ \omega_A)_{MiP}$.

2.1.2 Initial rise

The *initial rise* (also known as *initial lowering*) can be defined as an obligatory F0 rise at the beginning of a Minor Phrase. In (2a), where each word constitutes a Minor Phrase, each word starts with an initial rise (Figure 1, left). In (2b), where there is only one Minor Phrase, a clear F0 rise is found only at the beginning of the entire phrase. In the latter case, there is a high plateau after the initial F0 rise, which

3 The sample pitch contours in this chapter are created from data collected in various production experiments conducted by the author. They have been modified for expository purposes (parts of the contours affected by microprosody have been removed, and the contours have been smoothed).

continues until the accentual F0 fall on the accented word *wa'in o* 'wine ACC'.[4] Together with accent culminativity, the initial rise defines the Minor Phrase.

2.1.3 Downstep

Downstep is pitch range compression triggered by lexical pitch accents. In Figure 1 (left), the pitch range is narrowed down after the first accentual fall. As a result, the F0 peaks of the following words are realized much lower than that of the first word.

McCawley analyzed this F0 compression as the conversion of accentual high (H) tones into mid (M) tones, and defined the Major Phrase as the domain of this pitch reduction process. McCawley claimed that the pitch accent reduction applies uniformly to all non-initial H-tones within a Major Phrase. The first accent within a Major Phrase is realized with a H-tone, whereas all other pitch accent H-tones are converted to M-tones.

The pitch reduction after a lexical pitch accent was later given a different interpretation by other researchers, and came to be called *downstep* (or *catathesis*). Poser (1984), Pierrehumbert and Beckman (1988) and Kubozono (1988, 1993) all assumed, following McCawley, that the Major Phrase is the domain of downstep. The effect of downstep is canceled at the end of a Major Phrase, and *pitch reset* takes place at the beginning of the following Major Phrase.

Poser (1984), Pierrehumbert and Beckman (1988) and Kubozono (1988, 1989, 1993) also showed experimentally that the downstep effect appears on every tonal target in a cumulative fashion within a Major Phrase, contra McCawley's analysis.[5] Cumulative effects of downstep within Major Phrases and pitch reset at the beginning of a new Major Phrase are shown schematically in Figure 2 (see also Igarashi, this volume, for more discussion on downstep).[6]

In sum, Minor Phrase and Major Phrase are defined in terms of the three prosodic phenomena: accent culminativity, initial rise, and downstep.

(3) a. Minor Phrase: Domain of accent culminativity and initial rise
    b. Major Phrase: Domain of downstep

4 A small F0 rise at the beginning of the third word in Figure 1 (right) may potentially be analyzed as an initial rise. Then the phrasing would be $(\omega_U\ \omega_U)_{MiP}(\omega_A)_{MiP}$.

5 Selkirk and Tateishi (1991) support McCawley's (1968) view of non-cumulative downstep. See below.

6 Poser, Pierrehumbert and Beckman, and Kubozono have distinguished two different types of F0 downward trend. One is downstep, which is currently under discussion. The other is *declination*, a global, time-dependent F0 downtrend that takes place irrespective of prosodic structures. Figure 2 does not represent the latter effect.

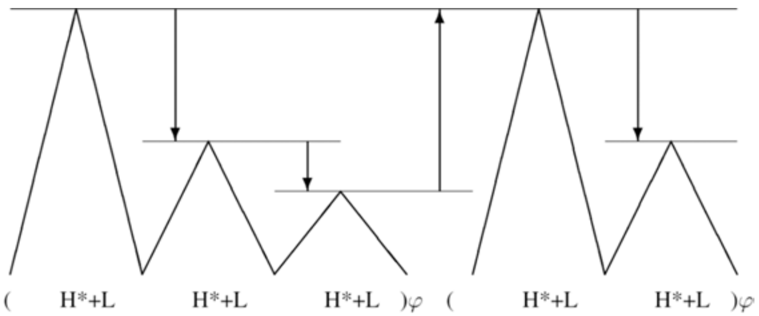


Figure 2: Schematic illustration of cumulative downstep and pitch reset (adapted from Ishihara 2011b: 1872).

2.1.4 Unifying Minor and Major Phrases

From the discussion above, the categorical distinction of Minor and Major Phrase seems well motivated for Japanese. From a cross-linguistic point of view, however, this distinction has proven difficult to motivate, as most of the world's languages lend themselves to analyses without any corresponding distinction. At the same time, there are cases where more than two levels seem to be needed (see below). The formulation of a fixed set of prosodic categories always involves a dilemma between the pursuit of explanatory adequacy and that of descriptive adequacy. In order to build a restrictive model of grammar, the number of categories should be kept to a minimum. In order to describe detailed empirical findings, the need for fine-grained level distinctions becomes stronger.

Ito and Mester (2007, 2012, 2013) circumvent this dilemma by appealing to two aspects of their theory. First, *recursivity* is integrated as part of prosodic structure.[7] A PPhrase may, for example, dominate another PPhrase. By allowing prosodic recursion as part of the basic principle of the prosodic hierarchy, the model can provide a sufficient number of prosodic levels to capture prosodic phenomena while keeping the number of prosodic categories cross-linguistically constant.

Second, they introduce two prosodic subcategories, which are given *relational* definitions, namely *maximal projections* and *minimal projections*, as illustrated in Figure 3. A prosodic category κ (e.g., PPhrase) that is not dominated by κ is a *maximal* projection of κ, while a prosodic category that is not dominating κ is a *minimal* projection of κ. By using relational terms to distinguish subcategories, the model allows a limited number of subcategories within each distinctive category.

7 See, among others, Ladd (1986, 1988), Gussenhoven (1991, 2005) and Selkirk (1996) for evidence for recursion in prosody.

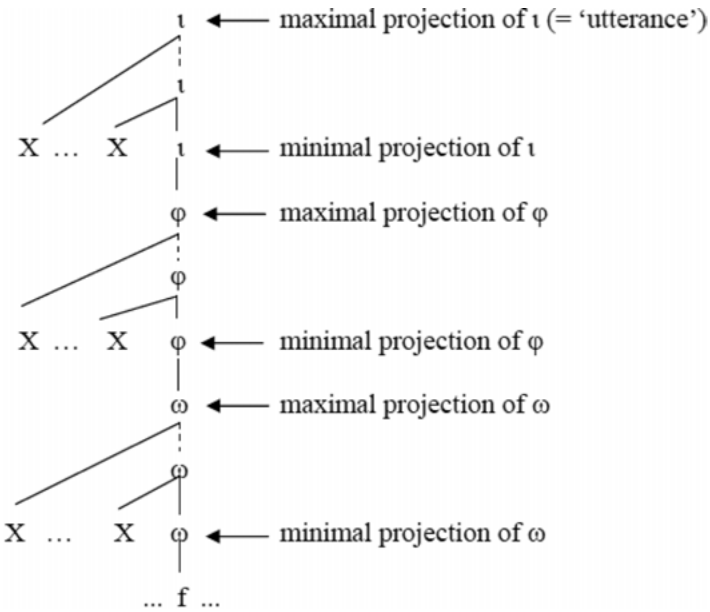


Figure 3: Prosodic recursion and relational definition of subcategories (Ito and Mester 2012: 288).

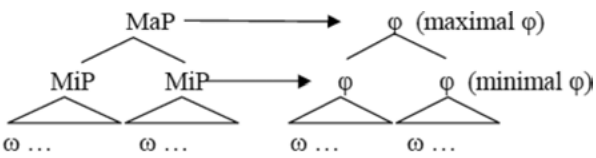


Figure 4: Redefinition of Minor and Major Phrases as recursive projections of PPhrase (Ito and Mester 2012: 290).

Ito and Mester (2012) claim that Minor and Major Phrases can be redefined as projections of the same category (PPhrase). They point out that the initial rise (previously used to define Minor Phrases) and downstep (previously used to define Major Phrases) can in fact be considered to apply to *all* instances of PPhrases, as in Figure 4. Since the left boundary of a Major Phrase always coincides with the left boundary of a Minor Phrase, the initial rise takes place at the beginning of every Minor *and* Major Phrase. Then the Minor/Major distinction is not necessary to account for initial rises. There is also no problem in assuming that downstep takes place not only within Major Phrases, but also within Minor Phrases, because downstep can only apply vacuously within the latter domain. Due to accent culminativity, Minor Phrases contain maximally one lexical pitch accent. Even if we assume that

downstep takes place after a single accent in a Minor Phrase, its effect would be canceled at the end of that Minor Phrase, and hence does not affect the realization of any following tonal targets. In other words, the effect of downstep within a Minor Phrase is never visible due to accent culminativity. Then the Minor/Major Phrase distinction is no longer needed to capture the domain of downstep.

Accent culminativity, too, can be accounted for in Ito and Mester's model. They claim that accent culminativity is an exclusive property of minimal PPhrases (i.e., the lowest level of PPhrase projections). By using the relational term "minimal projection", there is no need to postulate another distinctive prosodic category to account for the idea from previous models that accent culminativity is an exclusive property of Minor Phrases. After the unification of Minor and Major Phrases, the prosodic domains of the three prosodic cues can be redefined as follows.

(4) a. PPhrase (φ): Domain of initial rise and downstep
    b. Minimal PPhrase (φ_{min}): Domain of accent culminativity

2.2 PClause

In the previous subsection, we established tonal evidence for two levels of prosodic domains, Minor and Major Phrase, and their unification into a single category, PPhrase. A similar argument has been made for the level of the PClause. The *Utterance* and the *Intonational Phrase* (or *Intonation Phrase*) have been proposed as separate prosodic categories in the prosodic hierarchy (Nespor and Vogel 1986). For Japanese, however, Pierrehumbert and Beckman (1988) claimed that there is no empirical evidence for the Intonational Phrase, and hence do not adopt two distinct levels. Pierrehumbert and Beckman's analysis was adopted in the ToBI transcription systems for Japanese, J_ToBI (Venditti 2005) and X-JToBI (Maekawa et al. 2002; Venditti, Maekawa, and Beckman 2008; Igarashi, this volume). The Japanese ToBI systems, however, use the term "Intonation Phrase" to refer to what Pierrehumbert and Beckman (1988) referred to as "Intermediate Phrase" (= Major Phrase). See Table 1. Despite the terminological tweaking, they all agree that in Japanese there is only one prosodic level above the PPhrase.

Recently, however, Kawahara and Shinya (2008) presented experimental evidence that there are phonetic cues that indicate the existence of a prosodic category between the Utterance and the PPhrase, and claimed that Japanese also needs a level of Intonational Phrase that is distinct from the Utterance.

In Ito and Mester's (2012) model, the Utterance and the Intonational Phrase are subsumed under a single prosodic category, the PClause. They propose that the Utterance, the highest level of the prosodic hierarchy, should be redefined as the

maximal projection of the PClause, as indicated in Figure 3. In Ito and Mester's model, then, there is no need to keep the Utterance as a separate prosodic category.

3 Syntax–prosody mapping

Based on the phonological background established in the previous section, this section discusses theories of the *syntax–prosody mapping*, i.e., the correspondence between the syntactic structure and the prosodic structure. Syntax–prosody mapping is one of the fundamental components of the phonological theories of the syntax–phonology interface, as the syntactic structure of the sentence is one of the major factors that shape sentence prosody.

In the discussion of the syntax–prosody mapping, two questions need to be addressed. The first question is what type of syntactic information is relevant for the syntax–prosody mapping. It is usually assumed that only a limited amount of syntactic information is available to the phonological component, and that all other information is "invisible". A natural question to ask is, then, what exactly the syntactic information that is visible to the phonological component is. The second question is how syntactic factors interact with the non-syntactic factors which sometimes interfere with the syntax–prosody mapping, i.e., prosodic wellformedness constraints (e.g., accent culminativity and rhythmic effects) and information structure (e.g., focus and givenness, to be discussed in section 4). With these questions in mind, this section surveys theories of syntax–prosody mapping.

3.1 Early empirical findings

One of the earliest studies explicitly designed to investigate the relation between the syntactic structure and intonation in Japanese was conducted by Uyeno and her colleagues (Uyeno, Hayashibe, and Imai 1979; Uyeno, Hayashibe et al. 1980, 1981; Uyeno, Yamada et al. 1980). Uyeno, Hayashibe, and Imai (1979) found, through a series of production experiments, that there is a higher F0 peak at the beginning of conjoined clauses and relative clauses than at a clause-medial position. Syntactically ambiguous sentences like (5) are disambiguated by the difference in the pitch curve. If there is a clause boundary in front of a word (as in the second word *koronda* 'fell' in (5b)), the F0 peak of that word is realized higher than in the case where it appears clause-medially (as in (5a)). Uyeno, Yamada et al. (1980) also showed that this difference in pitch height is used in sentence comprehension as a cue for disambiguation of syntactic structures. Uyeno et al. (1981) extended the object of study to various declarative sentences containing subordinate clauses, and found that pauses also indicate clause-initial boundaries. In their studies, however, the

phonetic effects of syntactic boundaries found in these studies were not discussed in terms of phonological organization of prosodic constituents.

(5) a. *[[ototoi koronda] otona ga waratta]*
day.before.yesterday fell adult NOM laughed
'The adult who fell the day before yesterday laughed.'

b. *[ototoi [koronda] otona ga waratta]*
'The adult who fell laughed the day before yesterday.'
(Uyeno, Hayashibe, and Imai 1979: 184)

3.2 Phonetic superimposition model

A phonetically oriented, synthesis-based model developed by Fujisaki and his colleagues (Fujisaki and Sudo 1971; Fujisaki and Hirose 1984; Fujisaki 2004) does not assume any prosodic structure like the one discussed in section 2. Instead, each factor that affects the F0 contour is modeled as an independent mathematical command. According to acoustic and physiological restrictions, these commands are converted into functions of time (upward and downward movements over time), each of which represents an F0 movement predicted according to the commands. The actual output (F0 contour) is created by combining these functions. Ladd (1996/2008) calls this theory *superimposition theory*, because two (or potentially more) contours are superimposed onto each other to model the actual F0 contour.

Fujisaki proposed two types of commands, *phrase commands* and *accent commands*. The former express (global) effects expected from the syntactic phrasing, and the latter express (local) effects expected from pitch accents (i.e., initial rise and accentual fall). As shown by Uyeno and her colleagues, in Japanese, there is an upward F0 movement at the beginning of syntactic phrases or clauses (which corresponds to the pitch reset at PPhrase). In (6), which is from Fujisaki (2004), there is a syntactic phrase boundary between the subject and the predicate. Phrase commands provide upward impulses at the beginning of these syntactic phrases (expressed as upward arrows in Figure 5). The lowering effect at the end of a declarative sentence (which corresponds to the so-called *final lowering* of Pierrehumbert and Beckman 1988) is also expressed as a phrase command (expressed as a downward arrow).

In addition to the upward and downward movements expressed as phrase commands, there are effects on the F0 contour from pitch accents. The material surrounding a lexical pitch accent is grouped into a single prosodic unit (which corresponds to a Minor Phrase). In (6), four accentual units can be detected: (*ao'i*) (*aoi no e' wa*) (*yama no ue no ie' ni*) (*a'ru*). In each unit, there is an F0 rise at the beginning (= initial rise), and an accentual F0 fall after the lexical pitch accent. These effects are expressed as accent commands (expressed as boxes of different sizes in Figure 5).

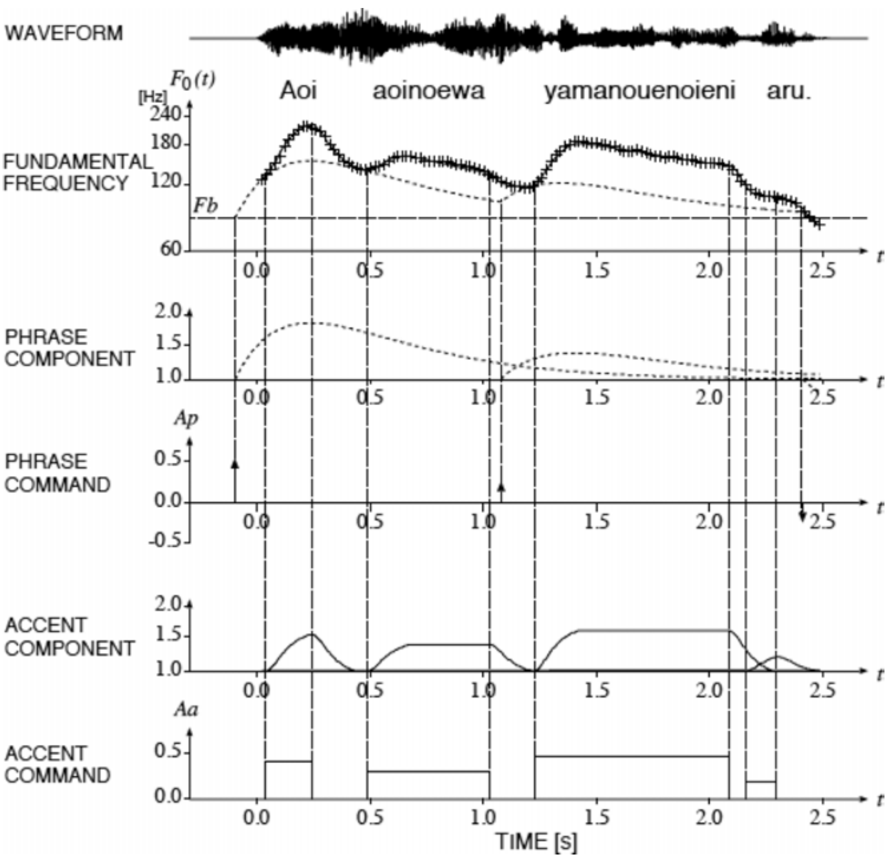


Figure 5: Schematic illustration of Fujisaki's model (from Fujisaki 2004).

(6)	*[$_{Subj}$*	*Ao'i*	*aoi*	*no*	*e'*	*wa]*	*[$_{Pred}$*	*yama*	*no*	*ue*	*no*
		blue	hollyhock	GEN	picture	TOP		hill	GEN	top	GEN

ie' *ni* *a'ru].*
house LOC exists
'The picture of the blue hollyhock is in a house on top of the hill.'

Phrase and accent commands are computed independently in the *phrase* and *accent components*, respectively, to create independent contours, based on the physiological and physical mechanisms of the larynx. As a result, each component produces an independent contour. The final output, i.e., the expected F0 contour, is created as a result of superimposition of the two functional curves from the phrase and the accent components, as illustrated in Figure 5.

Unlike the phonological theories that assume a hierarchical prosodic structure (cf. section 2), the phonetic superimposition model lacks an intermediate phonological representation that reflects effects of both accents and syntax simultaneously.

The model also seems to imply that lexical pitch accents (accent commands) have more local effects, while syntactic phrasing has more global effects. However, some effects of lexical pitch accents may in fact extend to larger domains than assumed in this model (see sections 3.3.3 and 4.2.2 for relevant discussion). See also Ladd (2008: 23–31) for a critical review of the superimposition models, both Fujisaki's and others'.

3.3 End-based and branching-based theories

3.3.1 End-based theory

Following claims by Selkirk (1986, 1996) and Chen (1987), an influential theory of syntax–prosody mapping has emerged which is called the *end-based theory.* This theory relies on the notion of *edge-alignment*, and claims that the syntactic information relevant for the syntax–prosody mapping is edges (i.e., boundaries) of syntactic maximal projections (XPs).

For Japanese, Selkirk and Tateishi (1991) claimed that the left edge of syntactic XPs is mapped as the left edge of prosodic PPhrases. Based on recordings of sentences with different syntactic structures, like those in (7)/Figure 6, they examined the correlation between syntactic XP boundaries and prosodic PPhrase boundaries.

(7) a. Left-branching subject noun phrase

$[_{VP}$ $[_{NP}$ $[_{NP}$ $[_{NP}$	*Ao'yama*	*no]*	*Yama'guchi*	*no]*	*ani'yome*	*ga]*
	Aoyama	GEN	Yamaguchi	GEN	sister-in-law	NOM

inai]
not.there
'We can't find the sister-in-law of Mr. Yamaguchi from Aoyama.'

b. Right-branching subject noun phrase
$[_{VP}$ $[_{NP}$ $[_{NP}$ Ao'yama no] $[_{NP}$ $[_{NP}$ Yama'guchi no] ani'yome ga]] inai]
'We can't find Mr. Yamaguchi's sister-in-law from Aoyama.'
(Selkirk and Tateishi 1991: 523)

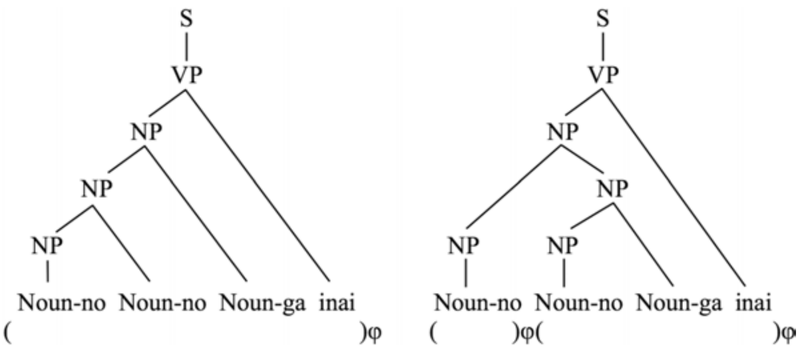


Figure 6: The syntactic structures of (7) and the PPhrase boundaries predicted by Selkirk and Tateishi's (1991: 531) analysis.

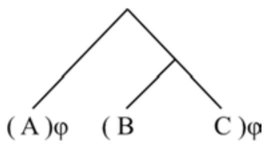
Selkirk and Tateishi (1991) identified PPhrase boundaries based on downstep and reset. In (7a), where the left edges of all the XPs coincide at the beginning of the sentence, the sentence shows downstep from the first noun to the second one. In (7b), where an XP (in this case NP) boundary intervenes between the first noun and the second, pitch reset is observed on the second noun. As discussed in section 2.1.3, pitch reset after downstep indicates the beginning of a new PPhrase (= the domain of downstep) and thereby the end of the preceding one.

This analysis establishes a strong correlation between syntactic XP boundaries and prosodic PPhrase boundaries.[8] The end-based analysis has become one of the major standard theories of the syntax–prosody mapping, and has been integrated into the Optimality Theoretic (Prince and Smolensky 2004) analyses (see section 3.5).

There are a few important aspects of the end-based theories to be noted. A first point is that in this analysis of Japanese, only the left edge of XPs is mapped onto prosody. Right edges are automatically inserted in front of every left edge (under the assumption that there is no recursion of PPhrases). A second point is that the correspondence between the syntactic boundaries and the prosodic boundaries is stated only in one direction. A typical end-based mapping principle states that syntactic boundaries correspond to prosodic boundaries, but not vice versa. Due to these two aspects (mapping applies to only one side of XPs, and only from syntax to prosody), the end-based analysis does not require a strong, one-to-one correspondence between syntactic and prosodic constituents, but a somewhat loose correspondence.

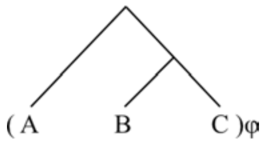
A third point is that the predictions made by the end-based theory depend on the syntactic analysis of the given structure, especially in terms of the *labels* assigned to syntactic nodes. For example, a right-branching structure like (8) will be phrased differently depending on the syntactic label of the intermediate branching node. If the intermediate node is analyzed as a maximal projection (XP) in the X-bar theory (Chomsky 1970), as in (8a), the theory predicts a PPhrase boundary between A and B. If it is analyzed as a non-maximal, bar-level projection (X'), as in (8b), the entire syntactic phrase will be mapped as a single PPhrase, because non-maximal projections are ignored. In the latter case, the mapping result is parallel to that of the left-branching structure in (9). In other words, the analysis predicts that the difference in the syntactic branching structure will be neutralized in the prosodic representation. As we will see below, however, this prediction is not tenable (at least in a straightforward way).

(8) a. XP-level right-branching: $[_{XP}$ A $[_{\mathbf{XP}}$ B C]] → (A)φ (B C)φ

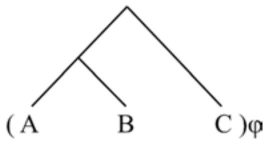


8 Selkirk and Tateishi (1988, 1991) ascribed the initial observation of this correspondence to Terada (1986).

b. X'-level right-branching: $[_{XP}\ A\ [_{\mathbf{X'}}\ B\ C]] \rightarrow (\ A\ B\ C\ )\varphi$



(9) XP/X'-level left-branching: $[_{XP}\ [_{XP/X'}\ A\ B]\ C] \rightarrow (\ A\ B\ C\ )\varphi$



3.3.2 Branching-based theories

While the end-based theory refers to specific labels in syntactic representations (namely, maximal projections) and their edges, the *branching-based analysis* uses syntactic branching as the cue for the syntax–prosody mapping. Furthermore, the branching-based analysis takes the syntactic effect to be cumulative, in contrast to Selkirk and Tateishi's analysis. The more syntactic brackets appear at a certain boundary, the stronger the corresponding prosodic effect becomes.

Kubozono (1988, 1989, 1993) investigated the amount of downstep in right-branching structures like (8) and left-branching structures like (9). In examining the amount of downstep, he used the so-called *paradigmatic* methodology. As in (10) and (11), for each syntactic structure, he compared two conditions, one with an accented word ω_A before the target word (<u>underlined</u>), and one with an unaccented word ω_U before the target word. Since only accented words trigger downstep, the difference in F0 height between the two conditions reveals the effect of downstep.

(10) a. Left-branching: $[[\boldsymbol{\omega}_{\mathbf{A}}\ \underline{\omega_A}]\ \omega_A]$

*[[**a'warena***	*<u>mi'nari</u>*	*<u>no</u>]*	*o'yako]*
poor	appearance	GEN	parent.and.child

b. Left-branching: $[[\boldsymbol{\omega}_{\mathbf{U}}\ \underline{\omega_A}]\ \omega_A]$

*[[**kiiroi***	*<u>ya'ne</u>*	*<u>no</u>]*	*ie'ie]*
yellow	roof	GEN	houses

(11) a. Right-branching: $[\boldsymbol{\omega}_{\mathbf{A}}\ [\underline{\omega_A}\ \omega_A]]$

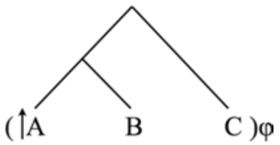
*[**kowa'i***	*[<u>me'</u>*	*<u>no</u>*	*ya'mai]]*
fearful	eye	GEN	disease

b. Right-branching: [$\boldsymbol{\omega_U}$ [$\underline{\omega_A}$ ω_A]]
*[**kiiroi** [<u>me'n</u> <u>no</u> ori'mono]]*
yellow cotton GEN fabric

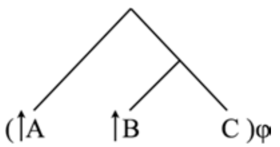
The results showed that both the left-branching and the right-branching structures exhibit an effect of downstep on the second word. Based on the assumption that a PPhrase boundary cancels the effect of downstep and triggers a (complete) pitch register reset, he concluded that there is no PPhrase boundary in either case, i.e., both structures have the same prosodic phrasing.

The results also showed, however, that the F0 peak of the second word is realized significantly higher in the right-branching structure (8) than in the left-branching structure (9). Kubozono claimed that there is an additional F0 boosting effect only in the right-branching structure, which he calls *metrical boost (MB)*. Metrical boost expands the pitch range at the left edge of each branching structure. This means that whenever there is a syntactic branching, the left node is subject to metrical boost. In his analysis, the two structures are analyzed as in (12), with metrical boost indicated with ↑.

(12) a. Left-branching: [[A B] C] → (↑A B C)φ



b. Right-branching: [A [B C]] → (↑A ↑B C)φ

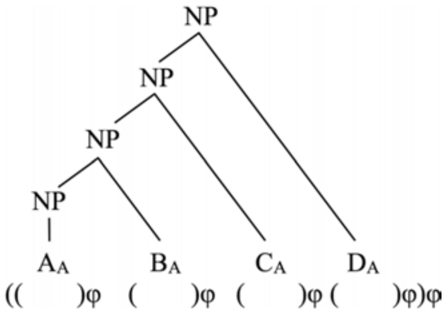


Furthermore, Kubozono found that the effect of metrical boost is cumulative. When more than one left bracket appears at the same place in a syntactic structure, a stronger metrical boost is found than when only one left bracket appears. Kubozono also claimed that right-branching structures are prosodically marked in Japanese, as they exhibit marked prosodic behaviors, and block various processes (e.g., accent phrase formation, rendaku, preaccentual boost).

Another of Kubozono's noteworthy findings is the *rhythmic effect*. In a left-branching structure containing four accented words ω_A like (13), metrical boost would be expected only at the beginning of the entire phrase, because all the left nodes coincide at the beginning of the phrase. (Accent culminativity requires that

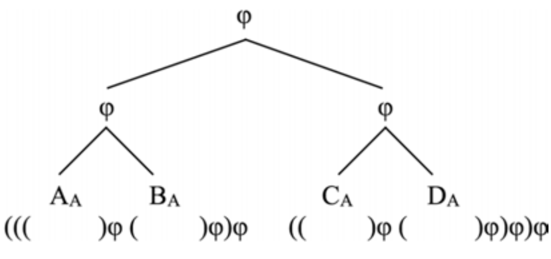
each accented word form a minimal PPhrase of its own.) As a result, successive downstep would be expected throughout the phrase.

(13) Uniformly left-branching structure with four accented words:
$[[[[\ \mathrm{A_A}\ ]\ \mathrm{B_A}\ ]\ \mathrm{C_A}\ ]\ \mathrm{D_A}\ ] \rightarrow ((\ \mathrm{A_A}\ )\varphi\ (\ \mathrm{B_A}\ )\varphi\ (\ \mathrm{C_A}\ )\varphi\ (\ \mathrm{D_A}\ )\varphi)\varphi$



Kubozono showed, however, that structures like (13) are realized with a raised F0 peak on the third word, although this word still exhibits some downstep effect. In order to account for this unexpectedly high F0 peak on the third word, Kubozono proposed another boosting effect, called *rhythmic boost*. As a result of rhythmic boost, the uniformly left-branching structure is prosodically phrased as two (intermediate) PPhrases containing two minimal PPhrases each, which are dominated by a single maximal PPhrase, as in (14). From this representation, downstep is expected from A to B, and from C to D, as well as from (A B) to (C D). Shinya, Selkirk, and Kawahara (2004) corroborate Kubozono's results.[9]

(14) Rhythmic effect: $[[[[\ \mathrm{A_A}\ ]\ \mathrm{B_A}\ ]\ \mathrm{C_A}\ ]\ \mathrm{D_A}\ ] \rightarrow (((\ \mathrm{A_A}\ )\varphi\ (\ \mathrm{B_A}\ )\varphi)\varphi\ ((\ \mathrm{C_A}\ )\varphi\ (\ \mathrm{D_A}\ )\varphi)\varphi)\varphi$



The prosodic structure in (14) clearly deviates from that predicted by the mapping principles in (13). Rhythmic effects can be regarded as one of the *prosodic wellformedness conditions*, which impose restrictions on the prosodic representations independently of the syntactic structure.

---

9 The intermediate level in (14) was treated as a recursive Minor Phrase by Kubozono, and was called "Superordinate Minor Phrase (SMiP)" by Shinya, Selkirk, and Kawahara (2004).

In Kubozono's model, the tonal organization is represented as PPhrases (Minor and Major Phrases in his terms), while the effect of syntactic branching is modeled as an independent metrical process, namely, metrical boost. The actual phonetic output is a combination of these two effects. His paradigmatic experimental methodology provides a precise criterion for the effect of downstep, by which the presence/absence of PPhrase boundaries can be determined.

The idea of the branching-based analysis and the cumulative effect of syntactic boundaries is more prominently put forward in Tokizaki's (2006) theory (*Bare Mapping Theory*). He claimed that syntactic bracketing is directly mapped onto prosody as prosodic boundaries.[10] As a result, the number of syntactic brackets at a certain word boundary is reflected in the strength of the phonetic realization of that boundary, such as the F0 boosting effect. This theory predicts a fine-grained relative strength of prosodic boundaries within a sentence. It contrasts in this sense with the end-based theory, in which the prosodic boundary effects are expected only at XP-boundaries.

One notable aspect of Tokizaki's analysis is that he does not assume specific prosodic categories for prosodic boundaries. What matters instead is the relative strength between boundaries. He claimed that the number of boundaries may decline as the speech rate increases. The relative strength between boundaries, however, is maintained in the process of boundary reduction. At boundaries where many brackets coincide, more boundary marks will remain after the reduction, while boundaries where fewer brackets coincide may disappear in faster speech.

3.3.3 Comparing Selkirk and Tateishi's (1991) and Kubozono's (1993) approaches

Two methodological problems can be pointed out in Selkirk and Tateishi's experiment. The first problem concerns their *syntagmatic* method for detecting pitch reset. Selkirk and Tateishi only examined sentences with accented words. Whether pitch reset had taken place on a target word (e.g., the second word in (7a)/(7b)) could thereby be determined only in relation to the preceding word within the same sentence. If the target word showed a higher F0 peak than the preceding word, pitch reset was assumed to have taken place, on the (tacit) assumption that a downstepped F0 peak cannot be higher than the preceding peak, which triggered the downstep.

Their approach was criticized by the proponents of the paradigmatic approach. Kubozono (1989, 1993) showed that even when an F0 peak is higher than the preceding peak, it may be subject to downstep, because that peak may be raised due to independent boosting effects (such as metrical and rhythmic boosts). A pure downstep effect is visible only with a paradigmatic methodology.

10 Syntactic nodes dominating phonologically empty elements (empty categories, null functional heads, etc.) as well as non-branching XPs are ignored (Uechi 1998; Tokizaki 2006).

The second methodological problem in Selkirk and Tateishi's experiment is their method of structurally disambiguating the stimuli. In (7), the subject NP contains identical words in the two conditions. The meaning difference is only distinguished by the difference in the syntactic structure. In order to achieve proper disambiguation, the speaker must be fully aware of the structural difference between the two conditions. Under such circumstances, it is possible that the speakers deliberately make extra effort to disambiguate the two conditions by adding prosodic prominence to the disambiguating area. (See, for example, Hirotani 2005 for the influence of speakers' awareness of the syntactic/semantic contrast on the prosodic phrasing.) In other words, the results may be affected by the effect of (contrastive) focus. As will be shown in section 4, prosodic effects of focus need to be distinguished from those of the syntax–prosody mapping.

In fact, the effect of focus that may potentially have appeared in Selkirk and Tateishi's experiment may be the source of two puzzling contradictions in the claims regarding downstep between Selkirk and Tateishi (1991) on the one hand and Poser (1984), Pierrehumbert and Beckman (1988) and Kubozono (1993) on the other.

The first contradiction is the interpretation of the results from the right-branching structure discussed above. Selkirk and Tateishi claim that there is no downstep on the second phrase in the right-branching structure, while Kubozono claims that this phrase is downstepped (in addition to being subject to metrical boost). The difference in their interpretation of the data does not seem to come only from the methodological difference between the syntagmatic and paradigmatic approaches mentioned above. Selkirk and Tateishi's data does not seem to show (despite the lack of the baseline condition) any lowering effect on the second word, while Kubozono's data, which replicates the data from Poser (1984) and Pierrehumbert and Beckman (1988), does show a lowering effect in the same position.

Comparing their results with Poser's (1984), Selkirk and Tateishi (1991: 532) claim that the difference comes from the difference in the syntactic labels of the relevant branching nodes in the tested stimuli. The right-branching sentences used by Selkirk and Tateishi are composed of a sequence of three nouns [N-GEN [N-GEN N]], while those tested by the other researchers contain two adjectives and a head noun [Adj [Adj N]]. Selkirk and Tateishi assume that their examples involve two XPs $[_{NP}$ N-GEN $[_{NP}$ N-GEN N]], while the examples in Poser (1984) involve a single XP $[_{NP}$ Adj $[_{N'}$ Adj N]]. With this structural difference, the end-based theory predicts the presence of a PPhrase boundary (and pitch reset) only in Selkirk and Tateishi's data.

Selkirk and Tateishi's explanation, however, cannot explain the difference between the left- and right-branching structures found in Kubozono's (1988, 1993) data. Selkirk and Tateishi's analysis predicts, as shown in (8b) and (9), that the syntactic difference between the left- and right-branching structures will be neutralized at the X'-level. Kubozono has shown that this is not the case.

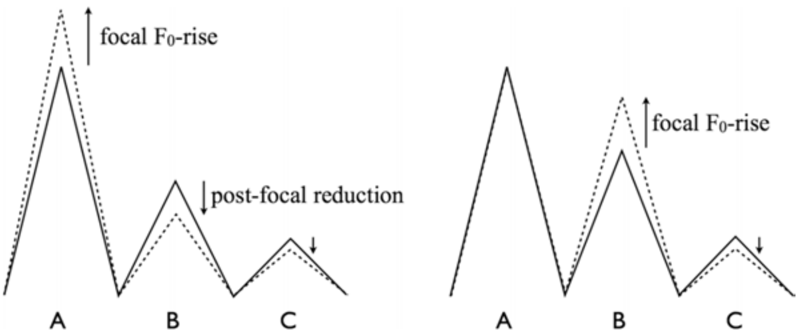


Figure 7: Expected focus effects (dashed lines) when disambiguating a left-branching (left) and a right-branching structure (right). Solid lines represent the contours without focus effects according to Kubozono's (1993) results.

The second point of contradiction between Selkirk and Tateishi on the one hand and Poser, Pierrehumbert and Beckman, and Kubozono on the other is the cumulative nature of downstep. In left-branching structures like (9), Selkirk and Tateishi's data show a large difference in F0 height between the first lexical pitch accent and the second one. The second pitch accent, however, does not seem to trigger further downstep. Based on this finding, they support McCawley's (1968) view (see section 2.1.3) that downstep takes place only between the first and second lexical pitch accents within a PPhrase. This claim contradicts the results reported by Poser (1984), Pierrehumbert and Beckman (1988) and Kubozono (1993), who all claim that downstep is cumulative within a PPhrase. Since Poser (1984) and Pierrehumbert and Beckman (1988) only investigate right-branching structures, their results cannot be compared with Selkirk and Tateishi's. Kubozono's (1993) data, however, does contain the left-branching structure. Nevertheless, his data shows a cumulative downstep effect. The experimental data in Ishihara (2011b) also corroborate Kubozono's results.

The two puzzling contradictions start to make sense once we hypothesize that Selkirk and Tateishi's data involve the focus effect from deliberate structural disambiguation. In order to disambiguate the left- and right-branching structures, a focus would be placed on the first word in the left-branching structure, and on the second word in the right-branching structure, as schematically shown in Figure 7 (see section 4 for a detailed discussion of focus prosody). In the right-branching structure, it is plausible that the F0 peak on the second word in Selkirk and Tateishi's data is raised due to focal F0 rise, which would obliterate the downstep effect. In the left-branching structure, the first word receives focal F0 rise and the following words undergo post-focal reduction. As a result, the second word is realized lower than expected from a pure downstep effect, masking the cumulative downstep effect.

Although it seems superior to Selkirk and Tateishi's syntagmatic approach, Kubozono's paradigmatic approach also has its own problems. First, his theory does not make any prediction as to where non-minimal PPhrase (= Major Phrase) boundaries appear in a given phrase/sentence.[11] The paradigmatic methodology is a solid way to judge whether PPhrase boundaries exist at a given location, but it provides no clue as to where boundaries occur. Kubozono's model only predicts the location of expected metrical boost.

Another (potential) problem is the basic assumption adopted in the paradigmatic approach. Kubozono assumes, along with Poser (1984), Pierrehumbert and Beckman (1988) and others, that the PPhrase is the domain of downstep. Accordingly, in the paradigmatic methodology, the existence of a PPhrase boundary is confirmed only if there is a "complete" pitch reset from downstep. Results from recent experimental studies seem to suggest, however, that such a criterion may be too strict, because no complete reset has been found even in cases where a PPhrase boundary is typically believed to exist. For example, most theories of the syntax–prosody mapping predict that a sentence of the form [$_{NP}$ Noun GEN Noun NOM] [$_{VP}$ [$_{NP}$ Noun ACC] Verb] yields two PPhrases, one containing the subject NP, and the other containing the VP (Noun GEN Noun NOM)φ (Noun ACC Verb)φ. In recent studies (e.g., Kubozono 2007), no complete reset was found if the subject NP contains lexical pitch accents. If complete reset is not found even in such a case, it is unclear whether complete pitch reset exists at all within an entire sentence.

A possible reinterpretation of the "incomplete" reset is to assume that the domain of downstep is larger than the PPhrase (i.e., PClause), or, both PPhrase *and* PClause. Then complete pitch reset would never be expected within a PClause. In such an analysis, however, the defining characteristics of PPhrase become obscure, because PPhrase is no longer the (unique) domain of downstep. A more radical possibility would be to abolish the PPhrase/PClause distinction, and explain all instances of initial rise as metrical boost, as in Tokizaki's (2006) model. Kubozono's model is in principle compatible with such an analysis. Further study is needed to better understand the relation between PPhrases and downstep/pitch reset.

3.4 Phase-based theories

The previous section reviewed the end-based and the branching-based theories, which differ in terms of the type of syntactic information relevant for the mapping. The next two subsections survey two approaches which differ in terms of the architecture of the grammar. *Phase-based* theories are derivational (i.e., they assume serial computations) whereas *Optimality Theoretic* theories are representational (i.e., they assume parallel computations).

---

11 For minimal PPhrase (= Minor Phrase) boundaries, see Chapters 1–3 of Kubozono (1993).

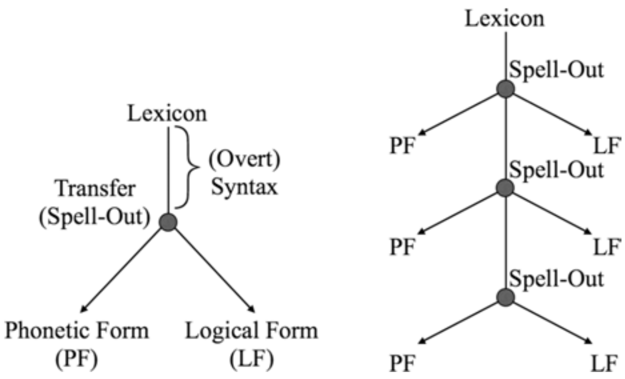


Figure 8: The “inverted Y” model (left) and the Multiple Spell-Out model (right).

Phase-based theories introduce the notions of *phases* and *Multiple Spell-Out* from the Minimalist Program (Chomsky 1995a, et seq.; Uriagereka 1999) to the theory of syntax–prosody mapping. In contrast to the traditional “inverted Y” model in generative syntax in Figure 8 (left), in the phase-based model, the syntactic structure is transferred to the two interface components PF and LF multiple times in the course of a single syntactic derivation, each time transferring a “smaller chunk” of a syntactic object to PF, as in Figure 8 (right). The domain of syntactic computations that create one of these “smaller chunks” is called a *phase*. Two syntactic projections are considered to be phases in the standard theories: CP and (transitive) *v*P (Chomsky 2000, 2001). At each phase, a subpart of the derivation (or the entire phase, depending on the theory) will be transferred to PF and LF. The domain that is transferred to PF/LF is called the *Spell-Out domain*. Spell-Out domains are often assumed to be the complement of the phase heads, i.e., VP for the *v*P phase, and TP for the CP phase (Chomsky 2000, 2001), though other proposals also exist (e.g., Fox and Pesetsky 2005; Shiobara 2010).

Chomsky originally proposed the notion of phase in order to derive the syntactic locality effects from the model. The phase theory, however, may also have phonological implications. Because of multiple Spell-Out, more frequent interaction between syntax and prosody can potentially be expected in a phase-based model than in a single Spell-Out model. The phase-based model also implies that there are syntactic operations after a first Spell-Out, which may potentially influence later syntactic operations. Such influences from phonology to syntax are not predicted by the single Spell-Out model. Furthermore, if a correspondence between prosodic domains and Spell-Out domains could be confirmed, this would provide empirical support for the phase theory.

Some researchers have therefore started exploring the implications of the phase-based syntactic theory for prosody (Legate 2003; Dobashi 2003; Ishihara 2003, 2007;

Kahnemuyipour 2009; Adger 2007; Kratzer and Selkirk 2007; Pak 2008; Sato 2009; Shiobara 2009, 2010; Revithiadou and Spyropoulos 2009, among others).

Dobashi (2003), adopting the label-free syntactic model of Collins (2002), proposed the phonological component creates prosodic structure based on Spell-Out domains and other prosodic wellformedness restrictions and parameterizations.

Ishihara (2003) proposed that prosody is computed derivationally, phase by phase, based on the cyclic and recursive nature of focus prosody in Japanese *wh*-constructions and the correspondence between the domain of focus prosody and the syntactic Spell-Out domains. Later, Ishihara (2007) pushed the idea of phase-based prosody further, and claimed that each Spell-Out domain is mapped to a PPhrase in the phonological component. That is, there is a one-to-one mapping of syntactic constituents (Spell-Out domains) to prosodic constituents (PPhrases).

Sato (2009, 2012) also proposed a phase-based prosodic model, in which prosody is computed within each phase and produced at its Spell-Out. He first pointed out that the claim made by Kahnemuyipour (2009) and Kratzer and Selkirk (2007), based on Persian and German data, respectively, makes the wrong predictions for Japanese. Sato (2012) proposed that language-specific parameterizations are needed to explain the placement of nuclear stress cross-linguistically.

Shiobara (2010: 10) also proposed a phase-based theory, in which the point of Spell-Out in a syntactic derivation is not determined by designated syntactic nodes (*v*Ps and CPs), but by correspondence between syntactic and prosodic categories. Any syntactic object may be spelled out if a corresponding prosodic constituent (according to Shiobara, the PPhrase for Japanese, and the PClause for English) is available.

There are various unsolved theoretical questions regarding the phase-based theories. First of all, the assumptions about which syntactic categories should count as a phase and a Spell-Out domain differ from analysis to analysis. It is therefore difficult to compare their theoretical predictions. A second issue is how to explain the interaction between the syntax–prosody mapping and non-syntactic factors, as many of the phase-based theories do not explicate how purely phonological effects such as accent culminativity and rhythmic effects on phrasing should be dealt with. Third, since the mechanism of multiple Spell-Out is arguably language-universal, prosodic variation needs to be explained in terms of language-specific parameterizations. Different researchers have proposed different types of parameterizations (e.g., Sato 2012 and Shiobara 2010). It remains to be seen if a unified account of parameterization can be achieved.

3.5 Optimality Theoretic analyses

In recent years, approaches based on the framework of Optimality Theory (OT hereafter) have been widely adopted in the study of the syntax–phonology interface in

various languages (Selkirk 1996, 2000, 2006; Truckenbrodt 1999; Samek-Lodovici 2005; Féry and Samek-Lodovici 2006; Prieto 2005; Feldhausen 2010; Myrberg 2013; Elfner 2012, among others) including Japanese (Truckenbrodt 1995; Ito and Mester 2003, 2013; Sugahara 2003; Selkirk 2009; Ishihara 2011b, among others). One of the major advantages of OT-based approaches in the study of the syntax–phonology interface is that they can treat constraints from different modules of the grammar (e.g., syntax, phonology, discourse) in a parallel manner, and express their interactions without being restricted by derivational steps (unlike the phase-based theories). In other words, OT is a useful framework to simulate how factors from different grammatical modules interact with each other. Another advantage of the theory is that it can express typological differences among languages/dialects with the same set of constraints by changing their rankings.

In standard OT approaches, the prosodic realization of an input syntactic structure is explained as a result of interactions between two types of constraints, namely, *interface constraints* and *prosodic wellformedness (markedness) constraints*, as in (15). Interface constraints express various types of relations that hold between two modules of the grammar, for example, the relation between syntactic and prosodic structures (syntax–prosody mapping constraints) or the relation between information structure and syntactic and/or prosodic structure (information structure–prosody mapping constraints, to be discussed in section 4). Interface constraints interact with prosodic wellformedness constraints, which impose restrictions on the realization of prosodic structure, for example, constraints on the size of prosodic constituents, on the location of prosodic heads/prominences, and so on. (See Truckenbrodt 2007 for an overview of the syntax–prosody interface in the OT framework.)

(15) OT constraints in the syntax–phonology interface
  a. Interface constraints:
    i. Syntax–prosody mapping
    ii. Information structure–prosody mapping (section 4)
  b. Prosodic wellformedness (markedness) constraints
    i. The Strict Layer Hypothesis (Exhaustivity, Nonrecursivity)
    ii. Constituent size (Binarity)
    iii. Accent culminativity

3.5.1 Syntax–prosody mapping constraints (interface constraints)

Two types of syntax–prosody mapping constraints have been proposed in the literature. The first type is the ALIGNMENT constraints (Selkirk 1996, 2000; Truckenbrodt 1995), which are the OT-version of the end-based theory discussed in section 3.3. The edge-based mapping conditions of the form “the left/right edge of X in syntax

corresponds to the left/right edge of Y in prosody" are translated into a set of ALIGNMENT constraints (Selkirk 1996), based on the idea of so-called *Generalized Alignment* (McCarthy and Prince 1993). In the case of Japanese, the left edge alignment for XPs and PPhrases can be formulated as in (16) (Selkirk and Tateishi 1991; Truckenbrodt 1995).[12]

(16) ALIGN(XP, L, φ, L) ("ALIGN-XP" for short)
Align the left edge of every XP with the left edge of a PPhrase.

ALIGN-XP carries over the basic properties of the end-based theory to OT. First, it requires a correspondence at only one side of syntactic/prosodic constituents. (The right edge alignment would be stated as a separate constraint, which is presumably ranked much lower than the left edge alignment in Japanese.) Second, it only requires a uni-directional correspondence relation between syntactic and prosodic boundaries. Each XP left boundary must coincide with a PPhrase left boundary, but not vice versa. A prosodic boundary not corresponding to a syntactic boundary does not cause violation of ALIGN-XP.[13] Another property of ALIGN-XP is that it is only sensitive to the maximal projections of *lexical* projections (N, A, V, P) and not those of *functional* projections (D, Agr, etc.) (Selkirk 1996; Truckenbrodt 1995, 1999; see Truckenbrodt 2007 for illustration).

The second type of syntax–prosody mapping constraints is MATCH constraints, proposed by Selkirk (2009, 2011). Selkirk claimed that prosodic categories above the rhythmic levels (i.e., above the foot) should be defined exclusively according to the syntactic categories (*The Syntax–Prosody Mapping Hypothesis*, *SPMH*). MATCH constraints require that syntactic categories (words, phrases, and clauses) be mapped to their corresponding prosodic counterparts (PWords, PPhrases, and PClauses) in the prosodic representation, as in (17). (See also Ito and Mester's theory discussed in section 2.)

(17) a. MATCHCLAUSE
A clause in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ι, in phonological representation.

b. MATCHPHRASE
A phrase in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it φ, in phonological representation.

12 The location of prosodic prominence (or heads of prosodic constituents) is also restricted by a set of ALIGNMENT constraints. A prosodic constituent of a specific level is usually defined as either left-headed or right-headed. See Truckenbrodt (1995, 2007).

13 Some researchers (e.g., Cheng and Downing 2009) postulate the reverse version of the mapping constraints (prosody-to-syntax mapping) to restrict the existence of prosodic boundaries that do not match with prosodic boundaries, following the basic idea of generalized alignment (McCarthy and Prince 1993). See the Match theory below for such bi-directional constraints.

c. MATCHWORD
A word in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ω, in phonological representation.
(Selkirk 2011: 439)

Furthermore, Selkirk (2011: 451) proposed that in addition to the three syntax-to-prosody mapping constraints, there are corresponding prosody-to-syntax mapping constraints (P–S faithfulness constraints), i.e., the mapping is bi-directional (see note 13).

The MATCH constraints require a much tighter, exact correspondence between syntactic and prosodic categories than ALIGN constraints, as both edges of a syntactic constituent must match with both edges of a corresponding prosodic constituent. (Note that ALIGN-XP allows many-to-one correspondences. A single PPhrase boundary can satisfy the alignment requirement of multiple XP boundaries.) Note also that the MATCH constraints obligatorily derive recursive prosodic structures, because MATCH requires exact mirroring of the syntactic constituent structure, which is recursive by nature, on the prosodic constituent structure.

The main idea is that prosodic categories are direct reflections of syntactic categories. In this sense, it may be said that the Match theory resembles the phase-based theories in its core spirit. The basic concept of exact mapping contrasts with the partial mapping of ALIGNMENT constraints, because the latter only require partial correspondence between syntax and prosody. At the same time, MATCH constraints have a property that contrasts with the phase-based theories. They are expected to interact with (and as a result sometimes be overridden by) prosodic wellformedness constraints.

3.5.2 Non-syntactic effects and prosodic wellformedness constraints

Syntax–prosody mapping constraints, both ALIGN and MATCH, require that the prosody mirror the syntactic structure. In the actual prosodic realizations, however, such requirements may be overridden by prosodic wellformedness constraints, depending on the ranking of relevant constraints.

The so-called *Strict Layer Hypothesis* (Selkirk 1984, 1986; Nespor and Vogel 1986) prohibits certain configurations in the prosodic structure, such as level-skipping (e.g., a PClause dominating a PWord), and recursion (e.g., a PPhrase dominating a PPhrase), and hence derives prosodic structures with a limited depth, unlike syntactic structures, which in principle may have unlimited depth and may contain recursion. Selkirk (1996) decomposed this hypothesis into four independent constraints: LAYEREDNESS, HEADEDNESS, EXHAUSTIVITY, and NONRECURSIVITY. By formulating EXHAUSTIVITY and NONRECURSIVITY as violable constraints, it is expected that these constraints may be violated under certain conditions. (Selkirk 1996: 190

claimed that the other two constraints, LAYEREDNESS and HEADEDNESS, are universally inviolable.) For example, an interaction of NONRECURSIVITY and the mapping constraints results in either a recursive or a flat structure. If NONRECURSIVITY in (18) is ranked lower than MATCH, as in (19), the recursive syntactic bracketing is maintained in prosody, violating NONRECURSIVITY. If the ranking is reversed, a non-recursive structure as in candidate b is chosen, violating MATCH.

(18) NONRECURSIVITY
No C^i dominates C^j, j = i,
e.g., "No Ft dominates a Ft."
(Selkirk 1996: 190)

(19)

	$[_{YP}\ [_{XP}\ X\ ]\ Y\ ]$	MATCH	NONREC
☞	a. $((\ X\ )_\varphi\ Y\ )_\varphi$		*
	b. $(\ X\ )_\varphi(\ Y\ )_\varphi$	$*!_{YP}$	

Another group of prosodic wellformedness constraints concern the minimal and maximal size of prosodic constituents and are called BINARITY constraints. With these constraints, the size of a prosodic constituent is often restricted to "at least two X" or "at most two X".

(20) a. MINIMALBINARITY(C_I), MINBIN
Prosodic constituent of level C_I must dominate at least two prosodic constituents of level C_{I-1}.
b. MAXIMALBINARITY(C_I), MAXBIN
Prosodic constituent of level C_I may dominate at most two prosodic constituents of level C_{I-1}.
(Sugahara 2003: 12)

In Japanese, MINBIN and MAXBIN play a role in the minimal PPhrase (Minor Phrase) formation. A sequence of unaccented words tends to be grouped into PPhrases containing either two or three words (Selkirk and Tateishi 1988). According to MINBIN(φ), which requires that a PPhrase (φ) be minimally binary, a PPhrase dominating a single PWord is dispreferred. MAXBIN(φ), on the other hand, would exclude a PPhrase dominating three or more PWords.

According to Selkirk and Tateishi (1988), Japanese allows a minimal PPhrase to contain three unaccented PWords ($\omega_U\ \omega_U\ \omega_U$). This is because it is impossible to satisfy MINBIN and MAXBIN simultaneously when there are exactly three PWords to be parsed. If three PWords are phrased into two PPhrases, (ω ω)(ω) or (ω)(ω ω), MAXBIN will be satisfied while MINBIN will be violated. If three PWords are grouped into a single PPhrase (ω ω ω), MINBIN is satisfied while MAXBIN is violated. The fact

that the latter pattern was found in Selkirk and Tateishi's (1988) data shows that MINBIN is ranked higher than MAXBIN in Japanese, as illustrated in (21).

(21)

	ω_U ω_U ω_U	MINBIN	MAXBIN
	a. (ω)(ω)(ω)	*!**	
	b. (ω)(ω ω)	*!	
	c. (ω ω)(ω)	*!	
☞	d. (ω ω ω)		*

When there are four unaccented words in a sequence, the output is uniformly (ω ω)(ω ω), which satisfies both MINBIN and MAXBIN simultaneously.

(22)

	ω_U ω_U ω_U ω_U	MINBIN	MAXBIN
	a. (ω)(ω)(ω)(ω)	*!***	
	b. (ω)(ω ω)(ω)	*!*	
	c. (ω ω ω)(ω)	*!	*
☞	d. (ω ω)(ω ω)		

Accent culminativity (maximally one pitch accent per minimal PPhrase, see section 2.1.1) can be captured by another prosodic wellformedness constraint that is ranked higher than MINBIN. Here, we adopt Ito and Mester's (2013: 30) ACCENT-AS-HEAD in (23):

(23) ACCENT-AS-HEAD
Every accent is the head of a φ_{min}.
Assign one violation for each accent that is not the head of a minimal phonological phrase φ.

If two or more accented PWords form a single PPhrase, each non-head PWord incurs a violation of ACCENT-AS-HEAD, as shown in (24):

(24)

	ω_A ω_A ω_A ω_A	ACC-AS-HEAD	MINBIN	MAXBIN
☞	a. (ω)(ω)(ω)(ω)		****	
	b. (ω)(ω ω)(ω)	*!	**	
	c. (ω ω)(ω ω)	*!*		
	d. (ω ω ω)(ω)	*!*	*	*

Before closing this section, a short discussion on Match theory is in order. Although Match theory is conceptually appealing (because it is much more restrictive in its application than Alignment theory), it needs further empirical support to be proven to be superior to the Alignment theory.

In fact, some of the previously made claims about Japanese have empirical as well as theoretical problems. Here, one such case will be discussed briefly. It is about the interaction between MATCH and BINARITY constraints. Selkirk (2011: 469) mentioned in passing that the rhythmic effect reported by Kubozono (1993) and Shinya, Selkirk, and Kawahara (2004), discussed in section 3.3.2, can be explained by interaction between MATCH constraints and MAXBIN. As shown in (14), a uniformly left-branching structure with four accented words $[[[[\omega_A]\ \omega_A]\ \omega_A]\ \omega_A]$ is phrased as a PPhrase dominating two smaller PPhrases, each of which contains two PWords, $((\omega_A\ \omega_A)\varphi\ (\omega_A\ \omega_A)\varphi)\varphi$. (Here we are simplifying the representation by omitting the minimal PPhrases required by accent culminativity.) At first glance, her argument appears to be convincing. The exact formalization reveals, however, that the MATCH-based analysis cannot select a correct output. In (25), MAXBIN, which bans any PPhrase that contains more than two PWords, successfully excludes candidate a, in which the two outermost brackets (indexed c and d) contain three and four PWords, respectively. MAXBIN, however, also excludes the desired output, candidate b, because the outermost brackets in this candidate contain four PWords. (Note that both candidates a and b have purely binary branching structures, which is irrelevant for MAXBIN.) The winning candidate is c, which lacks the outermost brackets. With this representation, no downstep is predicted on the third word C, contrary to fact.

(25)

	$[[[A]_1\ B]_2\ C]_3\ D]_4$	MAXBIN	MATCH(XP, φ)	MATCH(φ, XP)
	a. $((((A)_a\ B)_b\ C)_c\ D)_d$	$*_c!\ *_d$		
	b. $((\ A\ B\ )_a\ (\ C\ D\ )_b\ )_c$	$*_d!$	$*_3$	$*_b$
☞	c. $(\ A\ B\ )_a\ (\ C\ D\ )_b$		$*_3\ *_4$	$*_b$

It should be noted that this is not a problem specific to the BINARITY constraint, but a more general problem of MATCH. Once some prosodic wellformedness constraint outranks MATCH, MATCH constraints may lose their ability to create prosodic recursion. In such cases, we are left with a non-recursive, flat structure.

Another, more general note is that there are (at least) two methodological difficulties in confirming predictions of MATCH constraints. First, most phonetic cues for prosodic constituents reported in the literature (e.g., initial rise and downstep for the Japanese PPhrase) mark only one edge of the constituents. In order to confirm the presence of both edges predicted by MATCH constraints, two independent cues are

often needed.[14] Second, since MATCH constraints require a tighter correspondence between syntactic and prosodic constituents than ALIGN constraints, it is predicted that the former lead to more conflicts with prosodic wellformedness constraints than the latter. As a result, prosodic representations predicted by MATCH constraints are often overridden by prosodic wellformedness constraints, which makes it difficult to find empirical evidence for the tight syntax–prosody correspondence that MATCH constraints require.

For example, the prosody of embedded clauses illustrates the point. In the case of coordinated clauses, which have already been investigated by many researchers (Ladd 1986, 1988; Féry and Truckenbrodt 2005; Kawahara and Shinya 2008), each clause corresponds to a PClause, as predicted by MATCHCLAUSE. According to the MATCH constraints in (17), by contrast, a clausal complement within a VP, e.g., $[_{\mathrm{VP}}$ *think* $[_{CP}$ *that* ...]], would correspond to a PClause, while the VP containing it would correspond to a PPhrase. The resulting structure, (*think* {*that* ...}$\iota)\varphi$, clearly violates HEADEDNESS of the Strict Layer Hypothesis. Given that HEADEDNESS is an inviolable constraint (Selkirk 1996: 190), this phrasing would never be realized as it is. How such structures are excluded, and the right prosodic structure is derived through constraint ranking, still needs to be studied.

4 Information structure–prosody mapping

In the previous section, theories of the syntax–prosody mapping were reviewed. It was also shown that the syntax–prosody mapping is affected by prosodic wellformedness. This section discusses the interaction of the syntax–prosody mapping principles and another major factor that influences prosody, namely *information structure*, especially with respect to the notion of *focus*. In this chapter, the special prosody triggered by focus will be called *focus prosody.*

4.1 Focus, prominence, and emphasis

The definition of the term *focus* varies among different theoretical backgrounds. Furthermore, it is often used together with other terms such as “prominence” and “emphasis”. It is therefore important to clarify how these terms are distinguished here.

14 For Japanese, *Boundary Pitch Movements* (*BPMs*, section 4.2.1) may play the role of the right-edge maker. However, they appear only optionally, and they seem to mark PWords instead of PPhrases (see Igarashi, this volume, for discussion of BPMs).

First, this chapter adopts the following definition of focus from Krifka (2008: 247), which captures the central idea of the *alternative semantics theory* of focus (Rooth 1985, 1992).

(26) Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.

This means that only those elements that trigger a set of alternatives in the relevant discourse are taken to be focus. Discourse-new material (i.e., what has also been called *informational/presentational focus*: É. Kiss 1998; Selkirk 2002) does not qualify as being focused.

Second, it is assumed that each "prominence" denotes the head of some prosodic constituent, and that it is organized in a metrical grid representation (Liberman 1975; Liberman and Prince 1977; Nespor and Vogel 1986, 1989; Selkirk 1984; Truckenbrodt 1995). It has been claimed (Chomsky 1971; Jackendoff 1972; Truckenbrodt 1995) that focus requires (in many languages) that the highest prominence (within the scope of focus) be assigned somewhere in the focused constituent. Under this assumption, focus prosody is a result of manipulation of prosodic prominence triggered by focus. (Here, the prominence triggered by focus will be called *focal prominence*.) It is often assumed that focus is marked syntactically (e.g., *F-marking*: Jackendoff 1972), and that its scope is determined syntactically (e.g., by a focus operator ~: Rooth 1985, 1992). Then the main concern of the information structure–prosody mapping is to understand how (syntactically specified) focus, as well as its scope, is realized by prosodic prominence and interacts with syntax–prosody mapping.

Lastly, "emphasis" and "focus" will be distinguished from a semantic perspective, and only the prosodic effects related to the latter will be discussed in this chapter. According to the definition in (26), focus induces a set of alternatives in the discourse. Emphasis, on the other hand, does not necessarily have this grammatical function. For example, the adjective *dekai* 'big' can be emphasized by gemination of the second consonant, as in *dekkai* (Kori 1989b). This emphatic effect of consonant gemination, however, does not trigger any alternative to the adjective (e.g., other adjectives related to sizes such as *small*, *huge*, etc., or alternative degrees of "bigness", such as *not so big*, *slightly big*, *extremely big*, etc.).

4.2 Realizations of focus prosody in Japanese

Kori (1989a) showed that F0 exhibits much higher correlation with the presence/absence of semantic focus than intensity and duration. Despite the variation to be described below, focus prosody is generally characterized by two prosodic phenomena, which will be called *focal F0 rise* and *post-focal reduction*.

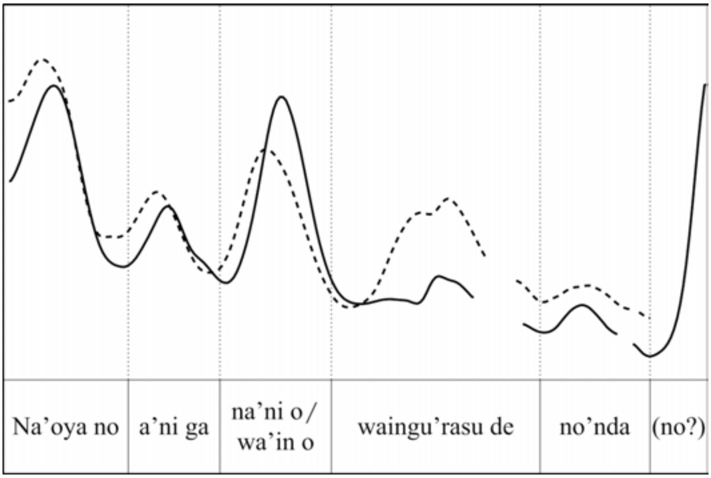


Figure 9: Sample pitch contours of a *wh*-question (27a) (solid line) and an all-new declarative counterpart (27b) (dashed line). A focal F0 rise is found on the third word *na'ni* 'what', followed by the post-focal reduction thereafter.

Focal F0 rise is a realization of focal prominence and is realized within the PWord containing a focused element.[15] *Post-focal reduction*[16] is a narrowing of the pitch excursion size in the post-focal area. In the case of (Tokyo) Japanese, focus prosody can be elicited in *wh*-questions like that in (27a) because *wh*-words in Japanese always behave like focused words, and therefore obligatorily trigger focus prosody (Maekawa 1991; Deguchi and Kitagawa 2002; Ishihara 2002, 2003). As shown in Figure 9, the *wh*-word *na'ni* 'what' in the *wh*-question (27a) shows a much higher F0 peak compared to the non-*wh*-counterpart in the declarative sentence in (27b). Furthermore, the pitch contour of the post-*wh*-area, *waingu'rasu de no'nda* 'drank with a wineglass' is more compressed than in the declarative counterpart. The pre-focal area, on the other hand, does not show much difference.[17]

(27) *Na'oya no a'ni ga [na'ni]$_F$/wa'in o waingu'rasu de*
Naoya GEN brother NOM what/wine ACC wineglass INST

no'nda (no?)
drank Q

a. *Wh*-question: 'What did Naoya's brother drink with a wineglass?'
b. Declarative: 'Naoya's brother drank wine with a wineglass.'

15 Focal F0 rise should be kept apart from initial rise at the beginning of PPhrases. See section 4.3.
16 This term was originally called "post-FOCUS reduction" by Sugahara (2003), in which "FOCUS" denotes contrastive focus.
17 Some pre-focal effects have also been reported. Hattori (1933) claimed that a focus on a word affects the realization of the pitch accents of preceding words. Maekawa (1997) reported that focus does not change the duration of the focused word, but reduces the duration of the entire utterance, by shortening the duration of pre-focal and post-focal areas.

The pitch contour in Figure 9 shows just one of the possible realizations of focal F0 rise and post-focal reduction. In fact, focal F0 rise may be realized in two different locations: it may appear on the focused word itself (early high pattern) or at the end of the PWord containing the focused word (late high pattern). Furthermore, post-focal reduction is also realized in two different ways, depending on whether the focused word and post-focal words are accented or not. The variation is summarized in (28):

(28) a. Focal F0 rise:
   i. Early high pattern: raising of the F0 peak on the focused word
   ii. Late high pattern: additional F0 rise on the PWord-final/penult mora
  b. Post-focal reduction:
   i. After pitch accents: compression with reduced pitch excursions
   ii. Unaccented area: high plateau with reduced pitch excursions

Below, variations of focal F0 rise and post-focal reduction will be illustrated in turn.

4.2.1 Two types of focal F0 rise

Oishi (1959) discussed two different realizations of (focal) prominence. The first type is raising of the F0 peak (pitch range expansion) of a focused word, as in Figure 9 above. Oishi called this the *early high pattern* (*maedaka-gata*). While the focal F0 rise can be detected for both accented and unaccented words, the amount of rise tends to be smaller with an unaccented word (Pierrehumbert and Beckman 1988).

In the second pattern, which Oishi called the *late high pattern* (*atodaka-gata*), there is an additional F0 peak on the last or penultimate mora of the focused PWord (= the content word plus a particle/postposition, if any). This peak is independent of the H-tone of the initial rise or lexical pitch accent of the focused word. (In (29) and (30), the location of the additional peak is indicated by capitals.) This means that in the case of an accented word, the realization of the late high pattern shows two F0 peaks.

(29) a. *sya.si'.n.KI.ga…* 'camera NOM'
  b. *ka'.re.RA.wa…* 'they TOP'
  (Oishi 1959: 87)

The late high pattern often allows variation in the location of the high peak. The F0 rise appears either on the last or on the penultimate mora.

(30) a. *hazime KAra vs. hazime kaRA* 'start' 'from'
b. *ii.tai DAke vs. ii.tai daKE* 'want.to.say' 'only/as much as'
(Oishi 1959: 90)

The late high pattern has been discussed by various researchers (Kawakami 1957; Kindaichi 1957; Muranaka and Hara 1994; Nagahara 1994; Oshima 2005; Kawahara and Shinya 2008; Venditti, Maekawa, and Beckman 2008, among others), and has been given different analyses. Kindaichi (1957) considered case-particles to have their own lexical pitch accents. Nagahara (1994) analyzed it as an initial rise after PPhrase boundary insertion at the left of focused case-markers (see section 4.3). In the ToBI frameworks, the late high pattern is analyzed as *Boundary Pitch Movements (BPMs)* (see Igarashi, this volume, for further discussion of BPMs).

4.2.2 Two realizations of post-focal reduction

In addition to the two types of focal F0 rises, post-focal reduction is also realized in two different ways, depending on the accentedness of the focused word and of the post-focal material. According to Ishihara (2011a), the post-focal area is realized with a compressed pitch contour (i.e., with a lower pitch range ceiling) following an F0 fall at a lexical pitch accent. In other words, accentual F0 fall triggers an extra compression in the post-focal area. For example, in Figure 9, in which all the words are accented, the post-focal area is realized with a more compressed contour compared to a non-focused counterpart.

If the focused word is unaccented, the pitch contour exhibits a high plateau following the focal F0 rise. Within this plateau, initial rises at PPhrase boundaries are observable, but their excursion size is smaller than in a non-focused counterpart, as exemplified in Figure 10. This suggests that the pitch range is compressed at a higher area, by raising the bottom of the pitch range.

(31) *Yamamoto wa [nani.go]$_F$/ainu.go no namae o*
Yamamoto TOP what.language/Ainu.language GEN name ACC
na'nnaku oboema'sita (ka?)
with.ease memorized Q
a. 'Names of what language did Yamamoto memorize with ease?'
b. 'Yamamoto memorized Ainu names with ease.'

Ishihara (2011a) further shows that the post-focal high plateau ends whenever there is a lexical pitch accent within the post-focal area, and the pitch contour after that shows the first type of post-focal reduction (i.e., accentual F0 fall followed by a compressed contour). For example, if the unaccented word following the unaccented *wh*-word in (31) *namae* 'name' is replaced with an accented word *myo'ozi* 'family

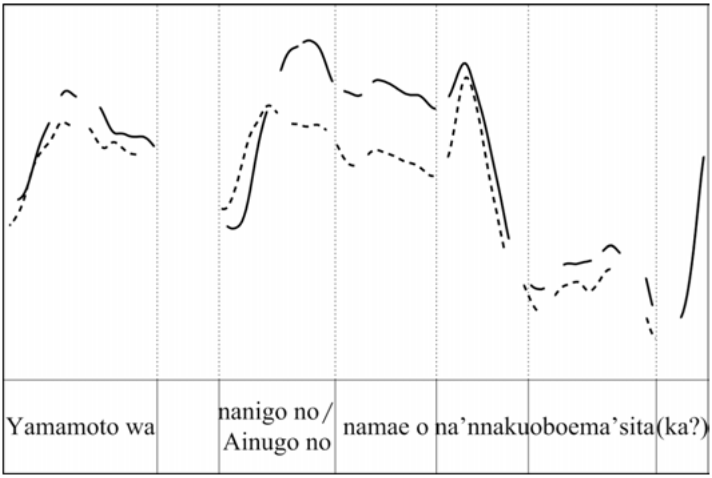


Figure 10: Sample pitch contours of a *wh*-question (31a) (solid line) and an all-new declarative (non-focus) counterpart (31b) (dashed line). The unaccented *wh*-word (*nanigo* 'what.language') shows a raised F0 peak followed by a high plateau, with a reduced pitch excursion (initial rise) at the beginning of the accented word (*na'nnaku* 'with ease').

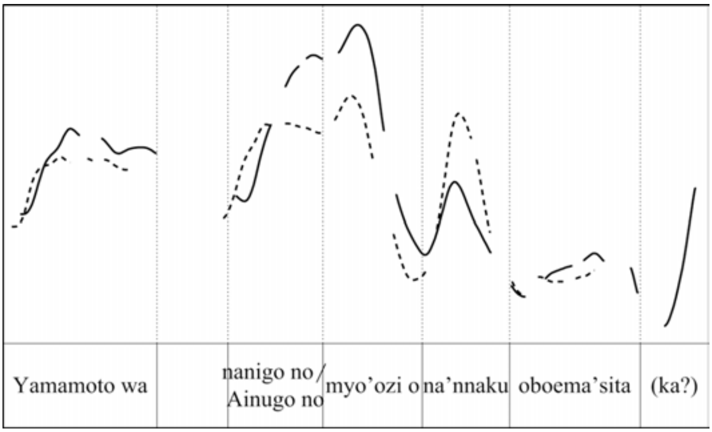


Figure 11: Sample pitch contours of a *wh*-question (solid line) and an all-new declarative (non-focus) counterpart (dashed line). The unaccented *wh*-word (*nanigo* 'what.language') shows a raised F0 peak followed by a high plateau until the following accented word (*myo'ozi* 'family.name'), and is followed by post-focal reduction thereafter.

name', there is a sharp F0 fall at this word, and the following word shows a post-accentual pitch range compression, as shown in Figure 11. This contour, as well as its further variants, is termed *emphasized unaccented accentual phrase* (*EUAP*) in the X-JToBI framework (Igarashi, Kikuchi, and Maekawa 2006: 361).

4.3 Two approaches to focus prosody

There have been two competing views on the phonological interpretation of Japanese focus prosody. The first view treats focus prosody as part of the prosodic structure, and derives it via manipulation of prosodic phrase boundaries (Pierrehumbert and Beckman 1988; Nagahara 1994; Truckenbrodt 1995; Selkirk 2000; Sugahara 2003). The other view considers the prosodic effects of focus independent of prosodic phrasing (Poser 1984; Shinya 1999; Kubozono 2007; Ishihara 2007). We will call the former *rephrasing analyses*, and the latter *non-rephrasing analyses*.

4.3.1 Rephrasing analyses

The basic idea of the rephrasing analyses is that the focal F0 rise and post-focal reduction are analyzed as results of PPhrase boundary insertion and deletion, respectively. PPhrase boundary insertion blocks downstep and triggers pitch reset. In contrast, PPhrase boundary deletion extends the domain of downstep, and blocks pitch reset.

For example, the "default" prosodic phrasing (i.e., the prosodic phrasing expected in an all-new discourse context) for the sentence in (32) is predicted to be like (32a), according to the syntax–prosody mapping principles (e.g., ALIGN-XP in (16)): a PPhrase boundary is expected between the subject DP and the predicate VP. With this prosodic phrasing, the second word *ani'yome ga* 'sister-in-law NOM' exhibits downstep from the preceding accented word *Ao'yama no* 'Aoyama GEN', followed by pitch reset at the following word *eri'maki o* 'scarf ACC'.

In contrast, if this sentence is produced in a discourse context where the second word *ani'yome ga* 'sister-in-law NOM' is focused, a focal F0 rise will appear on this word, and the following words exhibit post-focal reduction. In the rephrasing analysis, this is explained by insertion of a PPhrase boundary at the left edge of the focused word, and deletion of PPhrase boundaries after it, as shown in (32b). With this prosodic phrasing, downstep is no longer expected on the focused word, because the newly inserted PPhrase in front of it blocks downstep and instead triggers pitch reset. The following word, *eri'maki o* 'scarf ACC', on the other hand, will exhibit downstep, because pitch reset is no longer expected at this position. Under this view, the domain of focus prosody is seen as one single PPhrase.[18]

18 The rephrasing analysis can further be divided into two types. In the first analysis, which may be called the *direct rephrasing* analysis, focus directly affects prosodic phrasing, and inserts/deletes prosodic boundaries (e.g., Nagahara 1994; Selkirk 2000), as illustrated here. The second analysis may be called the *prominence-based rephrasing* analysis (Truckenbrodt 1995; Selkirk 2006). Focus affects the location of metrical prominence (stress), which in turn affects the prosodic phrasing (via head alignment constraints). In the latter analysis, the effect of focus on phrasing is indirect, mediated by prominence. See Ishihara (2011b) for a review of the prominence-based analyses.

(32) [$_{DP}$ *Ao'yama no* *[ani'yome* *ga]$_F$]* [$_{VP}$ *eri'maki* *o* *anda]*
Aoyama GEN sister-in-law NOM scarf ACC knitted
'Aoyama's sister-in-law knitted a scarf.'

a. (Ao'yama no ani'yome ga)φ (eri'maki o a'nda)φ

b. (Ao'yama no **)φ (** [ani'yome ga]$_F$ eri'maki o a'nda)φ

4.3.2 Non-rephrasing analyses

While the rephrasing analysis has been widely accepted, it has also been challenged by various empirical problems (Poser 1984; Shinya 1999; Kubozono 2007; Ishihara 2007; Féry and Ishihara 2010). In the non-rephrasing view, the prosodic effects of focus are independent of prosodic phrasing. In this line of analysis, focus prosody is a local modification of the pitch contour or pitch range, not of the prosodic phrasing. Since prosodic phrasing is not affected by focus under this analysis, focus prosody has often been considered to be a phonetic effect (an exception being Ishihara 2011b).

The proponents of the non-rephrasing analysis have presented various types of data that go against the rephrasing analysis. Regarding focal F0 rise, various empirical findings suggest that focus does not always coincide with a PPhrase boundary, i.e., focal F0 rise ≠ pitch reset at a PPhrase boundary. Poser (1984) first investigated the interaction of downstep and focal F0 rise in one of his datasets. He observed a downstep effect on the focused phrase, even when the F0 peak of the focused phrase is raised due to focus. This suggests that focus raises the F0 peak of the focused phrase, but does not block downstep triggered by the immediately preceding accented word. Extending Poser's study, Shinya (1999) investigated the focal F0 rise by using configurations where downstep takes place successively (left-branching structures), and showed that the existence of focus does not cancel the downstep effect completely, i.e., there is no complete pitch reset on the focused word. Kubozono (2007) reported that *wh*-phrases, which are always realized as focused phrases and hence trigger focal F0 rise, show the effect of downstep. Ishihara (2011b) showed that a focused phrase does not necessarily start with a boundary L-tone, which is a tonal indication of a PPhrase boundary.

Regarding post-focal reduction, data show that PPhrase boundaries are not necessarily removed in the post-focal area, i.e., post-focal reduction ≠ downstep. Sugahara (2003) claimed that post-focal reduction has the "dephrasing" effect (PPhrase boundary deletion) only when the post-focal material is given in the discourse. When the post-focal material is discourse-new, only a "non-structural" effect is found, in which all PPhrase boundaries are kept intact. Also, Ishihara (2007) claimed that post-focal reduction and downstep are independent phenomena. Within a post-focal domain, where the entire pitch register is compressed due to post-focal reduction, pitch reset can still be observed at the places where a PPhrase boundary is expected in the default prosodic phrasing.

Although there is ample empirical evidence that supports the non-rephrasing view, many proponents of the non-rephrasing view have kept it open as to how focus prosody should be theoretically explained. Either the focal F0 rise and post-focal reduction are considered purely phonetic effects (and hence should not be represented in a phonological representation), or some modifications to the rephrasing analysis should be made to capture the focus effects in the phonological representation. (See Ishihara 2011b for the latter analysis.)

4.4 Givenness

It has been claimed that discourse *givenness* has prosodic effects in various languages (Schwarzschild 1999; Baumann 2006; Féry and Samek-Lodovici 2006; Selkirk 2008; Katz and Selkirk 2011). When a linguistic expression is repeated in a discourse, i.e., when an expression is contextually given, the repeated expression is often prosodically reduced.[19] In English and other stress languages, givenness is typically realized as deaccentuation.

In Japanese, however, different observations have been made. Sugito (1985) claimed, based on her experimental data, that newness/givenness of information is not as clearly expressed prosodically in Japanese as in English.[20] Sugahara (2003), in contrast, showed experimentally that the realization of post-focal reduction is different depending on whether the post-focal material is discourse-new or given, as mentioned above. It has not yet been systematically investigated, to my knowledge, whether givenness has any prosodic effect outside the post-focal domain, and if so, in what way.

5 Studies from a syntactic perspective

In the introduction of this chapter, two perspectives toward the syntax–phonology interface were introduced. In sections 3 and 4, studies with phonological perspectives were discussed. This section briefly surveys several studies with a syntactic perspective. In these studies, syntactic phenomena, such as word order and semantic

19 Here the notion of givenness is taken as "repeated in the discourse". The exact definition of givenness, however, is still not an entirely settled issue in the literature. One of the well-known theories of givenness is the one by Schwarzschild (1999), in which givenness is defined as non-focus (i.e., non-F-marked), and is derived by semantic entailment using a semantic type shifting operation called "existential type shifting". Baumann (2006) and Baumann and Riester (2012) discuss different types of givenness. Gundel, Hedberg, and Zacharski (1989) propose the Givenness Hierarchy, in which five different cognitive statuses are hierarchically organized.

20 Sugito's results, however, appear to be affected by many other factors, such as the syntactic structures of the target sentences, the locations of the target words within the sentences, etc., which are not kept constant.

scope, are explained based on the prosodic properties of relevant syntactic elements or constructions.

5.1 Word order: scrambling

It has been shown in various languages that word order within a sentence may be affected by prosodic factors. A well-known example of this type of study addresses the relation between the nuclear stress of the sentence and word order. Reinhart (2006) and Neeleman and Reinhart (1998), for example, showed that the direct object in Dutch, which would receive the nuclear sentence stress in a canonical word order (Cinque 1993), moves across some other element (e.g., a sentence adverb) to be removed from the nuclear stress position, and as a result be interpreted outside of the focus of the sentence. Zubizarreta (1998) showed that in Romance languages like Spanish and Italian, the subject appears in a post-verbal position, where it receives the nuclear stress, when it needs to receive a focus interpretation.

5.1.1 Clause-internal scrambling

Scrambling is arguably the most intensively studied phenomenon in Japanese syntax (Harada 1977; Saito 1985; Miyagawa 1997; Bošković and Takahashi 1998; Ueyama 1998; Hiraiwa 2010, among many others). Several studies have claimed that prosody plays a role in this syntactic phenomenon of word order alternation. Ishihara (2001) claimed, adopting Reinhart's (2006) theory of Dutch scrambling, that Japanese clause-internal scrambling (Saito 1985; Tada 1993; Miyagawa 1997, 2003) has information-structural effects. It is claimed that in the default prosodic pattern (i.e., when no focus prosody is involved), the immediately preverbal phrase receives the nuclear prominence of the sentence,[21] and that any syntactic constituent containing this phrase can be interpreted as the new information of the sentence. Scrambling allows different combinations of phrases to be interpreted as part of new information. Under this view, scrambling results in a difference in the information structure, which is regulated by prosody. In a similar vein, Shiobara (2010) also claimed that clause-internal scrambling in Japanese is prosodically driven: the word order in which the focused element appears in the immediately preverbal position – the default sentence stress position – is preferred.

5.1.2 Long-distance scrambling

Sometimes in the syntax literature, so-called "PF-movements" are postulated when a movement operation does not have any semantic effect (i.e., the moved element is

21 Sato (2012) proposed a phase-based model (see section 3.4) to derive the nuclear stress position in Japanese.

interpreted at the original position, or the movement does not change the LF representation), or when it violates typical syntactic restrictions such as island conditions. In such cases, the reason for calling the operations "PF"-movement is not phonological in nature, and it often remains open whether there are any independent prosodic reasons to believe that such movements do take place in PF.

However, there are also studies in which the "PF" nature of such movements is attributed to certain prosodic properties of the moved element. For example, long-distance (henceforth, LD-)scrambling (e.g., Saito 1985, 1989, 1992) is known to exhibit certain prosodic restrictions. Koizumi (2000) observed that when two or more phrases in an embedded clause undergo (multiple) LD-scrambling, the moved constituents form a single prosodic constituent.[22] The acceptability judgment of the sentence becomes degraded without this prosodic phrasing.

Incidentally, multiple LD-scrambling exhibits not only this prosodic requirement, but also various peculiar syntactic behaviors (cf. Agbayani, Golston, and Ishii 2012 and the references therein). Based on this prosodic restriction on multiple LD-scrambling, some researchers have claimed that multiple LD-scrambling involves operations in PF. Fukui and Sakai (2003) claimed that multiply LD-scrambled phrases will be reanalyzed as a single constituent in PF (PF reanalysis) by an operation they call *Phrase-Level Merger*, along the lines of the *Morphological Merger* by Halle and Marantz (1993). Agbayani, Golston, and Ishii (2012) claimed that multiple LD-scrambling is in fact a single instance of a scrambling operation in PF applied to the prosodic constituent mentioned above. They claimed that PF scrambling is insensitive to syntactic conditions and restrictions, such as Condition C of the Binding Theory, locality constraints, etc.

Ishihara (2013) claimed that PClause boundaries play a crucial role in interpreting LD-scrambled phrases. He proposed a parsing principle, the *Principle of Argument Structure Parsing* (*PASP*), which states that XPs contained in a single PClause are preferably interpreted as clausemates (phrases that originate from the same syntactic clause). Assuming the Implicit Prosody Hypothesis (Fodor 1998, 2002), Ishihara claimed that the PASP applies even in silent reading, and influences the acceptability of sentences containing LD-scrambled phrases. He claimed that some of the phenomena discussed by Agbayani, Golston, and Ishii (2012) can be explained using the PASP, without assuming PF movement.

5.2 Semantic scope: wh-questions

Another area in which prosody seems to play a crucial role is the semantic scope of scope-taking elements. For example, Deguchi and Kitagawa (2002) and Ishihara (2002, 2003) found that focus prosody in *wh*-questions is not only obligatory (as shown in section 4.2), but also functions as a scope-marker.

22 Koizumi (2000) calls it an "intonational phrase", while Agbayani, Golston, and Ishii (2012) call it a "(recursive) phonological phrase".

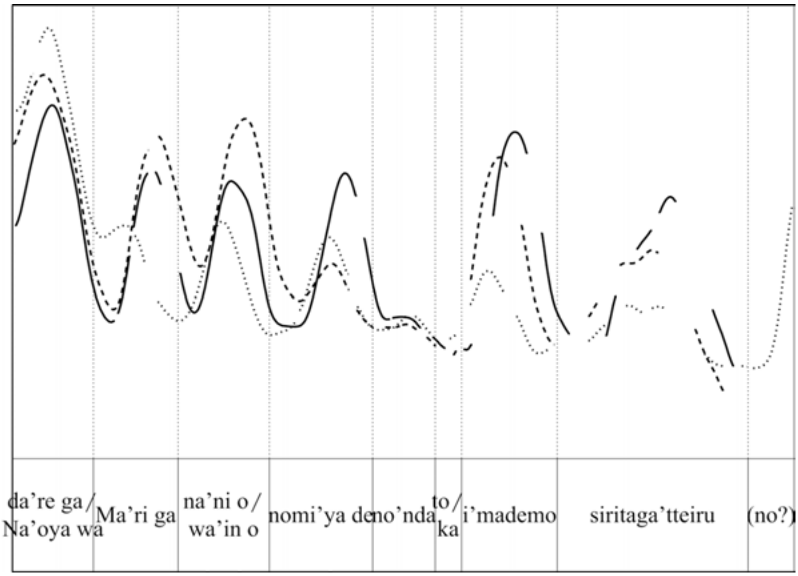


Figure 12: Sample pitch contours of a declarative sentence (33a) (solid line), a matrix *wh*-question (33b) (dotted line), and an indirect *wh*-question (33c) (dashed line). The matrix *wh*-question exhibits focal F0 rise on the first word (*da're* 'who'), followed by post-focal reduction until the end of the matrix clause. The indirect *wh*-question shows focal F0 rise on the third word (*na'ni* 'what'), followed by post-focal reduction until the end of the embedded clause (*ka* 'Q-particle'), and pitch reset at the penultimate word (*i'mademo* 'still').

The domain of focus prosody in *wh*-questions corresponds to the semantic scope of the *wh*-question, as illustrated in Figure 12. In a matrix *wh*-question like (33b) (the dotted line in Figure 12), where the semantic scope of the *wh*-question is the entire matrix clause, the *wh*-phrase is realized with a focal F0 rise, and the post-focal reduction (indicated with underlining) continues until the end of the matrix clause. In an indirect *wh*-question like (33c) (the dashed line in Figure 12), where the semantic scope of the *wh*-question is the embedded clause, the post-focal reduction continues only until the end of the embedded clause, and thereafter the pitch range is reset to normal (i.e., as high as in the declarative sentence, (33a), the solid line in Figure 12).[23] The same kind of prosodic scope-marking is also found in other constructions, e.g., the so-called *indeterminate constructions* (Kuroda 1965, 2013; Shimoyama 2001), as well as between negative polarity items (NPIs) and the associated negation (Tomioka 2007; Ishihara 2010).

23 In the Fukuoka dialect *wh*-questions exhibit a special prosody different from focus prosody (Hayata 1985; Kubo 1989; Smith 2005; Hwang 2011). The scope-marking function of this *wh*-prosody is parallel to that of focus prosody in Tokyo Japanese. A *wh*-prosody similar to that of Fukuoka Japanese may also appear in Tokyo Japanese, but only in indeterminate constructions (Ishihara 2003: 73; Kuroda 2013).

(33) a. Declarative sentence in the all-new condition: no focus prosody

Na'oya	*wa*	*[Ma'ri*	*ga*	*wa'in*	*o*	*nomi'ya*	*de*	*no'nda*	*to]*
Naoya	TOP	Mari	NOM	wine	ACC	bar	LOC	drank	that

i'mademo	*omo'tteiru*
still	think

'Naoya still thinks that Mari drank wine at the bar.'

b. Matrix wh-question: focus prosody until the end of the matrix clause

da're	*ga*	*[Ma'ri*	*ga*	*wa'in*	*o*	*nomi'ya*	*de*	*no'nda*	*to]*
who	NOM	Mari	NOM	wine	ACC	bar	LOC	drank	that

i'mademo	*omo'tteiru*	***no****?*
still	think	Q

'Who still thinks that Mari drank wine at the bar?'

c. Indirect wh-question: focus prosody within the embedded clause

Na'oya	*wa*	*[Ma'ri*	*ga*	***na'ni***	*o*	*nomi'ya*	*de*	*no'nda*	***ka****]*
Naoya	TOP	Mari	NOM	what	ACC	bar	LOC	drank	Q

i'mademo	*obo'eteiru*
still	remember

'Naoya still remembers what Mari drank at the bar.'

Since *wh*-scope is marked prosodically, a syntactically ambiguous sentence like (34) is disambiguated by appropriate focus prosody, as shown by Deguchi and Kitagawa (2002) and Ishihara (2002).[24]

(34)

Na'oya	*wa*	*[Ma'ri*	*ga*	*na'ni*	*o*	*nomi'ya*	*de*	*nonda*	*ka]*
Naoya	TOP	Mari	NOM	what	ACC	bar	LOC	drank	Q

Yu'mi	*ni*	*mora'sita*	*no?*
Yumi	DAT	divulged	Q

a. 'Did Naoya divulge to Yumi what$_i$ Mari drank t_i at the bar?'

b. (*?)[25] 'What$_i$ did Naoya divulge to Yumi whether Mari drank t_i at the bar?'

---

24 Hirotani (2005) conducted a series of psycholinguistic experiments on this type of construction and proposed the *Scope Prosody Correspondence* principle, which states that "[the] scope of a term X should not extend beyond the [PPhrase] containing X" (Hirotani 2005: 7). See Kitagawa and Hirose (2012) and Kitagawa, Tamaoka, and Tomioka (2013) for further investigation of focus prosody and its scope-marking property from psycholinguistic perspectives assuming the Implicit Prosody Hypothesis (Fodor 1998, 2002).

25 The acceptability of this interpretation varies among speakers, and generally, there is a bias toward the other reading. See Kitagawa and Fodor (2003) for more discussion.

If the prosodic *wh*-scope-marking is interfered with by other prosodic factors (e.g., another focus within the same sentence that requires an additional focal prominence), the appropriate reading becomes unavailable. This means that prosodic factors (such as obligatory focus prosody in *wh*-questions) sometimes interfere with the proper interpretation of a sentence, even though the sentence is syntactically sound. Ishihara (2002, 2003) showed that several phenomena discussed in the literature, such as (alleged) overt *wh*-questions in Japanese (Takahashi 1993), a missing reading in multiple *wh*-questions and indeterminate constructions (Shimoyama 2001), and the additional *wh*-effect (Kurata 1991; Saito 1994), can be explained prosodically, without postulating any ad-hoc syntactic conditions.

6 Conclusion and remaining issues

This chapter discussed various issues related to the syntax–phonology interface in Japanese, giving special attention to studies with a phonological perspective. They aim to establish a theory of prosody which explains the interaction of various factors that shape prosody, including syntax (syntax–prosody mapping), prosody (prosodic wellformedness), and information structure (information structure–prosody mapping). Some studies that take a syntactic perspective were also briefly reviewed. In these studies, syntactic phenomena such as word order and semantic scope-taking have been explained based on prosodic properties of the relevant constructions.

There are still many remaining questions to be investigated in the syntax–phonology interface. In regard to the prosodic hierarchy in Japanese (section 2), the Syntax–Prosody Mapping Hypothesis (Selkirk 2009, 2011; Ito and Mester 2007, 2012, 2013), which states that there are three and only three distinctive categories (PWord, PPhrase, and PClause) language-universally, has to be empirically examined. In particular, more study is needed for the PClause, the highest level of the hierarchy. Although Kawahara and Shinya (2008) presented evidence that clauses correspond to PClauses, they only investigated coordinated clauses. It still needs to be examined whether other types of syntactic clauses (embedded clauses, adjunct clauses, etc.) also exhibit prosodic cues for the PClause boundaries. Also, the categorical distinction between PPhrase and PClause needs to be motivated by further empirical evidence. Although the PPhrase has been assumed to be the domain of downstep, it still has to be examined whether the PClause also shows the downstep effect.

Regarding the syntax–prosody mapping (section 3), there are conflicting claims that need to be examined. As shown in section 3.3, the end-based theories and the branching-based theories make different predictions in certain configurations. The notions of pitch reset and metrical boost are very similar, but differently represented in the prosodic structure. The validity of the “syntagmatic” and “paradigmatic” methodologies also depends on the reexamination of the downstep domain, mentioned above. An example of not yet fully explored areas is the prosodic realization

of ditransitive VPs (VPs containing an indirect object, a direct object, and a verb). If the VP is analyzed as a single maximal projection [$_{VP}$ IO [$_{V'}$ DO V]], the end-based theory predicts a single PPhrase (IO DO V)φ, with downstep on DO and V. If a so-called "VP-shell", a recursive VP structure [$_{VP2}$ IO [$_{VP1}$ DO V_1] V_2], is assumed (Larson 1988; Chomsky 1995b; Ura 1999, among others), two PPhrases are expected (IO)φ (DO V)φ, with downstep only on V. (See Sato 2012 and Shiobara 2010 for discussion of ditransitive idioms.) Similar comparisons can be made for the two OT-based syntax–prosody mapping constraints, ALIGN-XP and MATCH-XP. Which of the two mapping constraints makes better predictions has to be carefully examined.

Concerning the information structure–prosody mapping (section 4), the issue related to the two analyses of focus prosody (rephrasing vs. non-rephrasing) has not been completely settled. In addition, the prosody of givenness needs further systematic investigation. The relation between boundary pitch movements (BPMs) and information structure is also yet another area to be further studied.

One area that this chapter has not touched on is the prosody of topic. Japanese has been recognized as an example of a language that has morphological topic marking (Kuno 1973). This does not mean, however, that prosody plays no role in expressing topicality. The so-called thematic and contrastive topics show different prosody (Nakanishi 2001).

Regarding the studies from the syntactic perspective (cf. section 5), a systematic model of syntax–prosody interaction is not yet fully established. In the framework of generative syntax, the flow of information is in principle unidirectional (syntax → phonology), although the multiple Spell-Out model has somewhat increased the possibility of frequent interactions between syntax and phonology. The influence of prosody in sentence processing, along the lines of the studies adopting the Implicit Prosody Hypothesis, may shed light on the prosody and acceptability judgments for "allegedly" syntactic conditions (for relevant discussion, see Hirose's chapter in the Psycholinguistics volume).

Acknowledgments

I would like to thank Haruo Kubozono, Yosuke Igarashi, and Sara Myrberg for useful comments and suggestions. Thanks also go to Yasunori Takahashi and Kirsten Brock for careful format and English corrections. All remaining errors are mine.

References

Adger, David. 2007. Stress and phasal syntax. *Linguistic Analysis* 33. 238–266.

Agbayani, Brian, Chris Golston and Toru Ishii. 2012. Prosodic Scrambling. In Ryan Bennett, Oliver Northrup, Anie Thompson and Matt Tucker (eds.), *Proceedings of the Fifth Formal Approaches*

to Japanese Linguistics Conference (MIT Working Papers in Linguistics 64), 15–24. Cambridge, MA: MIT.

Baumann, Stefan. 2006. *The intonation of givenness*. Tübingen: Niemeyer.

Baumann, Stefan and Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta and Pilar Prieto (eds.), *Prosody and meaning*, 119–162. Berlin: Mouton de Gruyter.

Beckman, Mary and Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bošković, Žejiko and Daiko Takahashi. 1998. Scrambling and last resort. *Linguistic Inquiry* 29(3). 347–366.

Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 2. 109–149.

Cheng, Lisa Lei-Shen and Laura Downing. 2009. Where's the topic in Zulu? *The Linguistic Review* 26. 207–238.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs and Peter Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Washington: Georgetown University Press.

Chomsky, Noam. 1971. Deep structure, surface structure and semantic interpretation. In Danny D. Steinberg and Leon A. Jakobovits (eds.), *Semantics, an interdisciplinary reader in linguistics, philosophy and psychology*, 183–216. Cambridge, UK: Cambridge University Press.

Chomsky, Noam. 1995a. Bare phrase structure. In Gert Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, 385–439. Oxford, UK: Blackwell.

Chomsky, Noam. 1995b. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels and Juan Uriagereka (eds.), *Step by step: In honor of Howard Lasnik*, 85–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24. 239–298.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein and T. Daniel Seely (eds.), *Derivation and explanation in the Minimalist Program*, 42–64. Oxford, UK: Blackwell.

Deguchi, Masanori and Yoshihisa Kitagawa. 2002. Prosody and Wh-questions. In Masako Hirotani (ed.), *Proceedings of the North Eastern Linguistic Society 32*, vol. 1, 73–92. Amherst, MA: GLSA.

Dobashi, Yoshihiko. 2003. *Phonological phrasing and syntactic derivation*. Ithaca: Cornell University dissertation.

Elfner, Emily Jane. 2012. *Syntax–prosody interaction in Irish*. Amherst, MA: University of Massachusetts dissertation.

É. Kiss, Katalin. 1998. Identification focus and information focus. *Language* 74. 245–341.

Feldhausen, Ingo. 2010. *Sentential form and prosodic structure of Catalan*. Amsterdam/Philadelphia: John Benjamins.

Féry, Caroline and Shinichiro Ishihara. 2010. How focus and givenness shape prosody. In Malte Zimmermann and Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 36–63. Oxford, UK: Oxford University Press.

Féry, Caroline and Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82(1). 131–150.

Féry, Caroline and Hubert Truckenbrodt. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59(2/3). 223–243.

Fodor, Janet Dean. 1998. Learning to parse? *Journal of Psycholinguistic Research* 27(2). 285–319.

Fodor, Janet Dean. 2002. Psycholinguistics cannot escape prosody. *Proceedings of Speech Prosody 2002*, 83–88.

Fox, Danny and David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31(1–2). 1–46.

Fujisaki, Hiroya. 2004. Information, prosody, and modeling: With emphasis on tonal features of speech. *Proceedings of Speech Prosody 2004.*

Fujisaki, Hiroya and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan* 5(4). 233–242.

Fujisaki, Hiroya and Hiroshi Sudo. 1971. A model for the generation of fundamental frequency contours of Japanese word accent. *Journal of the Acoustical Society of Japan* 27. 445–453.

Fukui, Naoki and Hiromu Sakai. 2003. The visibility guideline for functional categories: Verb raising in Japanese and related issues. *Lingua* 113. 321–375.

Gundel, Jeanette, Nancy Hedberg and Ron Zacharski. 1989. Givenness, implicature and demonstrative expressions in English discourse. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 25: Papers from the Parasession on Language in Context*, 89–103.

Gussenhoven, Carlos. 1991. The English rhythm rule as an accent deletion rule. *Phonology* 8. 1–35.

Gussenhoven, Carlos. 2005. Procliticized phonological phrases in English: Evidence from rhythm. *Studia Linguistica* 59. 174–193.

Halle, Morris and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale and Samuel J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Harada, S. I. 1977. Nihongo ni henkei wa hitsuyō da [There are transformations in Japanese]. *Gengo* 6(10–11). Reprinted in Naoki Fukui (ed.) 2000. *Syntax and meaning: S. I. Harada collected works in linguistics*, 545–566. Tokyo: Taishukan.

Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology.* Tokyo: Kaitakusha.

Hattori, Shiro. 1933. *Akusento to hōgen* [Accent and dialects]. Tokyo: Meiji Shoin.

Hayata, Teruhiro. 1985. Hakata hōgen no akusento/keitairon [The accent and morphology of the Hakata dialect]. Fukuoka: Kyushu University Press.

Hiraiwa, Ken. 2010. Scrambling to the edge. *Syntax* 13(2). 133–164.

Hirotani, Masako. 2005. *Prosody and LF: Processing Japanese wh-questions*. Amherst, MA: University of Massachusetts dissertation.

Hwang, Hyun Kyung. 2011. *Scope, prosody, and pitch accent: The prosodic marking of Wh-scope in two varieties of Japanese and South Kyeongsang Korean*. Ithaca: Cornell University dissertation.

Igarashi, Yosuke. this volume. Chapter 13: Intonation. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Igarashi, Yosuke, Hideaki Kikuchi and Kikuo Maekawa. 2006. Inritsu jōhō [Prosodic information]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kokuritsu Kokugo Kenkyūjo hōkoku 124: Nihongo hanashi kotoba kōpasu no kōchikuhō* [Construction of the Corpus of Spontaneous Japanese], 347–453. Tokyo: National Institute for Japanese Language.

Ishihara, Shinichiro. 2001. Stress, focus, and scrambling in Japanese. In Elena Guerzoni and Ora Matushansky (eds.), *A few from building E39: Papers in syntax, semantics and their interface* (MIT Working Papers in Linguistics 39), 151–185. Cambridge, MA: MIT.

Ishihara, Shinichiro. 2002. Invisible but audible wh-scope marking: Wh-constructions and deaccenting in Japanese. In Line Mikkelsen and Chris Potts (eds.), *Proceedings of the 21st West Coast Conference on Formal Linguistics* (WCCFL 21), 180–193. Somerville, MA: Cascadilla Press.

Ishihara, Shinichiro. 2003. *Intonation and interface conditions*. Cambridge, MA: MIT dissertation.

Ishihara, Shinichiro. 2007. Major Phrase, Focus Intonation, and Multiple Spell-Out (MaP, FI, MSO). *The Linguistic Review* 24. 137–167.

Ishihara, Shinichiro. 2010. Negative polarity item and focus intonation in Japanese. In Shoichi Iwasaki, Hajime Hoji, Patricia M. Clancy and Sung-Ock Sohn (eds.), *Japanese/Korean Linguistics 17*, 311–325. Stanford: CSLI.

Ishihara, Shinichiro. 2011a. Focus prosody in Tokyo Japanese wh-questions with lexically unaccented wh-phrases. *Proceedings of the 17th International Congress of Phonetic Science* (ICPhS XVII), 964–969.

Ishihara, Shinichiro. 2011b. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121(13). 1870–1889.

Ishihara, Shinichiro. 2013. Implicit prosody and acceptability: The case of long-distance scrambling in Japanese. In Kazuko Yatsushiro and Uli Sauerland (eds.), *Proceedings of the Sixth Formal Approaches to Japanese Linguistics* (MIT Working Papers in Linguistics 66), 57–72. Cambridge, MA: MIT.

Ito, Junko and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge, MA: MIT Press.

Ito, Junko and Armin Mester. 2007. Prosodic adjunction in Japanese compounds. In Yoichi Yamamoto and Masao Ochi (eds.), *Proceedings of the 4th Formal Approaches to Japanese Linguistics* (MIT Working Papers in Linguistics 55), 97–112. Cambridge, MA: MIT.

Ito, Junko and Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Browsky, Shigeto Kawahara, Takahito Shinya and Mariko Sugahara (eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk*, 280–303. London: Equinox Publishing.

Ito, Junko and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124(1). 20–40.

Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Kahnemuyipour, Arsalan. 2009. *The syntax of sentential stress*. Oxford, UK: Oxford University Press.

Katz, Jonah and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771–816.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto and Takahito Shinya. 2008. The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase and Utterance. *Phonetica* 65(1–2). 62–105.

Kawakami, Shin. 1957. Tōkyōgo no takuritsu kyōchō no onchō [Tonal prominence in Tokyo Japanese]. *Kokugo Kenkyū* 6. 21–31. Reprinted in Shin Kawakami 1995. *Nihongo akusento ronshū* [Papers on Japanese accent], 76–91. Tokyo: Kyūko Shoin.

Kindaichi, Haruhiko. 1957. Nihongo no akusento [Japanese accent]. In Etsutaro Iwabuchi, Ōki Hayashi, Hatsutaro Ōishi and Takesi Sibata (eds.), *Kōza gendai kokugogaku II: Kotoba no taikei* [Series in modern Japanese linguistics 2: The system of language], 67–94. Tokyo: Chikuma Shobō.

Kitagawa, Yoshihisa and Janet Dean Fodor. 2003. Default prosody explains neglected syntactic analysis in Japanese. In William McClure (ed.), *Japanese/Korean Linguistics 12*, 267–279. Stanford: CSLI Publications.

Kitagawa, Yoshihisa and Yuki Hirose. 2012. Appeals to prosody in Wh-interrogatives: Speakers' versus listeners' strategies. *Lingua* 122(6). 608–641.

Kitagawa, Yoshihisa, Katsuo Tamaoka and Satoshi Tomioka. 2013. Prosodic matters in intervention effects in Japanese: An experimental study. *Lingua* 124. 41–63.

Koizumi, Masatoshi. 2000. String vacuous overt verb raising. *Journal of East Asian Linguistics* 9. 227–285.

Kori, Shiro. 1989a. Fōkasu jitsugen ni okeru onsei no tsuyosa, jizoku jikan, F0 no yakuwari [The role of intensity, duration, F0 in focus realization]. *Onsei Gengo* 3. 29–38.

Kori, Shiro. 1989b. Kyōchō to intonēshon [Emphasis and intonation]. In Miyoko Sugito (ed.), *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō)* [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1], 316–342. Tokyo: Meiji Shoin.

Kratzer, Angelika and Elisabeth Selkirk. 2007. Phase theory and prosodic spell-out: The case of verbs. *The Linguistic Review* 24. 93–135.
Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276.
Kubo, Tomoyuki. 1989. Fukuoka-shi hōgen no dare, nani tō no gimonshi o fukumu bun no picchi patān [The pitch pattern of the sentences containing interrogative markers such as dare, nani, etc.]. *Kokugogaku* 156. 71–82.
Kubozono, Haruo. 1988. *The organisation of Japanese prosody*. Edinburgh: University of Edinburgh dissertation.
Kubozono, Haruo. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6. 39–67.
Kubozono, Haruo. 1993. *The organization of Japanese prosody*. Tokyo: Kurosio.
Kubozono, Haruo. 2007. Focus and intonation in Japanese: Does focus trigger pitch reset? In Shin-ichiro Ishihara (ed.), *Proceedings of the 2nd Workshop on Prosody, Syntax, and Information Structure* (Interdisciplinary Studies on Information Structure 9), 1–27. Potsdam: University of Potsdam.
Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.
Kurata, Kiyoshi. 1991. *The syntax of dependent elements*. Amherst, MA: University of Massachusetts dissertation.
Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. Cambridge, MA: MIT dissertation.
Kuroda, S.-Y. 2013. Prosody and the syntax of indeterminates. *Lingua* 124. 64–95.
Ladd, D. Robert. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.
Ladd, D. Robert. 1988. Declination 'reset' and the hierarchical organization of utterance. *Journal of the Acoustical Society of America* 84. 530–544.
Ladd, D. Robert. 1996/2008. Intonational phonology. Cambridge, UK: Cambridge University Press.
Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–391.
Legate, Julie Anne. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34(3). 506–516.
Liberman, Mark. 1975. *The intonational system of English*. Cambridge, MA: MIT dissertation.
Liberman, Mark and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.
Maekawa, Kikuo. 1991. Tōkyō hōgen gimon bun no intonēshon [The intonation of question sentences in Tokyo dialect]. *Proceedings of the 1991 PSJ Annual Convention*, The Phonetic Society of Japan, 42–47.
Maekawa, Kikuo. 1997. Effects of focus on duration and vowel formant frequency in Japanese. In Yoshinori Sagisaka, Nick Campbell and Noriko Higuchi (eds.), *Computing prosody: Computational models for processing spontaneous speech*, 129–153. New York: Springer-Verlag.
Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti. 2002. X-JToBI: An extended J-ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing* (ICSLP), Denver, Colorado, 1545–1548.
McCarthy, John J. and Alan S. Prince. 1993. Generalized alignment. *Yearbook of Morphology*. 79–153.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Miyagawa, Shigeru. 1997. Against optional scrambling. *Linguistic Inquiry* 28. 1–26.
Miyagawa, Shigeru. 2003. A-movement scrambling and options without optionality. In Simin Karimi (ed.), *Word order and scrambling*, 177–200. Oxford, UK: Blackwell.
Muranaka, Toshiko and Noriyo Hara. 1994. 'Joshi no takuritsu' no hindo to bunpō kinō oyobi onkyōteki tokuchō ni tsuite [Frequency, functions and acoustic features of prominent particles in Japanese discourse]. *Gengo Bunka Kenkyū* 1. 193–208.

Myrberg, Sara. 2013. Sisterhood in prosodic branching. *Phonology* 30(1). 73–124.
Nagahara, Hiroyuki. 1994. *Phonological phrasing in Japanese*. Los Angeles: University of California dissertation.
Nakanishi, Kimiko. 2001. Prosody and information structure in Japanese: A case study of topic marker *wa*. In Noriko M. Akatsuka and Susan Strauss (eds.), *Japanese/Korean Linguistics 10*. 434–447. Stanford: CSLI.
Neeleman, Ad and Tanja Reinhart. 1998. Scrambling and PF interface. In Mirium Butt and Wilhem Gueder (eds.), *The projection of arguments: Lexical and compositional factors*, 309–353. Stanford, CA: CSLI.
Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.
Nespor, Marina and Irene Vogel. 1989. On clashes and lapses. *Phonology* 6(1). 69–116.
Oishi, Hatsutaro. 1959. Purominensu ni tsuite: Tōkyōgo no kansatsu ni motozuku oboegaki [On prominence: Notes based upon the observation of Tokyo Japanese]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kotoba no kenkyū* 1, 87–102. Tokyo: NLRI.
Oshima, David Y. 2005. Boundary tones or prominent particles? Variation in Japanese focus-marking contours. *Berkeley Linguistics Society (BLS) 31*, 453–464.
Pak, Marjorie. 2008. *The postsyntactic derivation and its phonological reflex*. Philadelphia: University of Pennsylvania dissertation.
Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT dissertation.
Pierrehumbert, Janet and Mary Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.
Prieto, Pilar. 2005. Syntactic and eurhythmic constraints on phrasing decisions in Catalan. *Studia Linguistica* 59(2–3). 194–222.
Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford, UK: Blackwell.
Reinhart, Tanya. 2006. *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.
Revithiadou, Anthi and Vassilios Spyropoulos. 2009. A dynamic approach to the syntax-phonology interface: A case study from Greek. In Kleanthes K. Grohmann (ed.), *InterPhases: Phase-theoretic investigations of linguistic interfaces*, 202–233. Oxford, UK: Oxford University Press.
Rooth, Mats. 1985. *Association with Focus*. Amherst, MA: University of Massachusetts dissertation.
Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75–116.
Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. Cambridge, MA: MIT dissertation.
Saito, Mamoru. 1989. Scrambling as semantically vacuous A'-movement. In Mark R. Baltin and Anthony S. Krock (eds.), *Alternative conceptions of phrase structure*, 192–200. Chicago, IL: University of Chicago Press.
Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1. 69–118.
Saito, Mamoru. 1994. Additional-*wh* effects and the adjunction site theory. *Journal of East Asian Linguistics* 3. 195–240.
Samek-Lodovici, Vieri. 2005. Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23(3). 687–755.
Sato, Yosuke. 2009. Spelling-out prosodic domains: A multiple spell-out account. In Kleanthes K. Grohmann (ed.), *InterPhases: Phase-theoretic investigations of linguistic interfaces*, 234–259. Oxford, UK: Oxford University Press.

Sato, Yosuke. 2012. Phonological interpretation by phase: Sentential stress, domain encapsulation, and edge sensitivity. In Ángel J. Gallego (ed.), *Phases: Developing the framework*, 283–307. Berlin: Mouton de Gruyter.

Schwarzschild, Roger. 1999. GIVENness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141–177.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence prosody. *Phonology Yearbook* 3. 371–405.

Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Selkirk, Elisabeth. 2000. The interaction of constraints on prosodic phrasing. In Merle Horne (ed.), *Prosody: Theory and experiment*, 231–261. Dordrecht: Kluwer Academic Publishers.

Selkirk, Elisabeth. 2002. Contrastive FOCUS vs. presentational *focus*: Prosodic evidence from right node raising in English. In Bernard Bel and Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 643–646.

Selkirk, Elisabeth. 2006. Bengali intonation revisited: An optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing. In Chungmin Lee, Matthew Gordon and Daniel Büring (eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 215– 244. Dordrecht: Springer.

Selkirk, Elisabeth. 2008. Contrastive focus, givenness and the unmarked status of "Discourse-New". *Acta Linguistica Hungarica* 55(3–4). 331–346.

Selkirk, Elisabeth. 2009. On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu* 136. 35–74.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle and Alan Yu (eds.), *The handbook of phonological theory*, 2nd edition, 435–484. Oxford, UK: Blackwell.

Selkirk, Elisabeth and Koichi Tateishi. 1988. Constraints on minor phrase formation in Japanese. In Lynn MacLeod, Gary Larson and Diane Brentari (eds.), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 24-I: The General Session*, 316–336.

Selkirk, Elisabeth and Koichi Tateishi. 1991. Syntax and downstep in Japanese. In Carol Georgopoulos and Roberta Ishihara (eds.), *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda*, 519–543. Dordrecht: Kluwer Academic Publishers.

Shimoyama, Junko. 2001. *Wh-constructions in Japanese*. Amherst, MA: University of Massachusetts dissertation.

Shinya, Takahito. 1999. Eigo to nihongo ni okeru fōkasu ni yoru daunsuteppu no soshi to chōon undō no chōgō [The blocking of downstep by focus and articulatory overlap in English and Japanese]. *Proceedings of Sophia University Linguistic Society 14*, 35–51.

Shinya, Takahito, Elisabeth Selkirk and Shigeto Kawahara. 2004. Rhythmic boost and recursive minor phrase in Japanese. *Proceedings of the Second International Conference on Speech Prosody*, 183–186.

Shiobara, Kayono. 2009. A phonological view of phases. In Kleanthes K. Grohmann (ed.), *Inter-Phases: Phase-theoretic investigations of linguistic interfaces*, 182–201. Oxford, UK: Oxford University Press.

Shiobara, Kayono. 2010. Derivational linearization at the syntax–prosody interface. Tokyo: Hituzi Syobo.

Smith, Jennifer L. 2005. On the WH-question intonational domain in Fukuoka Japanese: Some implications for the syntax-prosody interface. In Shigeto Kawahara (ed.), *Papers on prosody* (University of Massachusetts Occasional Papers in Linguistics 30), 219–237. Amherst, MA: GLSA.

Sugahara, Mariko. 2003. *Downtrends and post-FOCUS intonation in Japanese*. Amherst, MA: University of Massachusetts dissertation.

Sugito, Miyoko. 1985. Nihongo to eigo ni okeru “shinjōhō” no hatsuwa no onkyōteki tokuchō [The acoustic characteristics of “new information” in Japanese and English]. In the Acoustical Society of Japan (ed.), *Onsei no kenkyū, vol. 21*, 335–348.

Tada, Hiroaki. 1993. *A / A-bar partition in derivation*. Cambridge, MA: MIT dissertation.

Takahashi, Daiko. 1993. Movement of wh-phrases in Japanese. *Natural Language and Linguistic Theory* 11. 655–678.

Terada, M. 1986. Minor phrasing in Japanese. Unpublished ms., University of Massachusetts, Amherst.

Tokizaki, Hisao. 2006. *Linearizing structure with silence: A minimalist theory of syntax-phonology interface*. Tsukuba: University of Tsukuba dissertation.

Tomioka, Satoshi. 2007. Pragmatics of LF intervention effects: Japanese and Korean Wh-interrogatives. *Journal of Pragmatics* 39(9). 1570–1590.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus, and prominence*. Cambridge, MA: MIT dissertation.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2). 219–255.

Truckenbrodt, Hubert. 2007. The syntax-phonology interface. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 435–456. Cambridge, UK: Cambridge University Press.

Uechi, Akihiko. 1998. *An interface approach to topic/focus structure*. Vancouver: University of British Columbia dissertation.

Ueyama, Ayumi. 1998. *Two types of dependency*. Los Angeles: University of Southern California dissertation.

Ura, Hiroyuki. 1999. Checking theory and dative subject constructions in Japanese and Korean. *Journal of East Asian Linguistics* 8(3). 223–254.

Uriagereka, Juan. 1999. Multiple spell-out. In Samuel David Epstein and Norbert Hornstein (eds.), *Working minimalism*, 251–282. Cambridge, MA: MIT Press.

Uyeno, Tazuko, Hideo Hayashibe and Kunihiko Imai. 1979. On pitch contours of declarative, complex sentences in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 13. 175–187.

Uyeno, Tazuko, Hideo Hayashibe, Kunihiko Imai, Hiroshi Imagawa and Shigeru Kiritani. 1980. Comprehension of relative clause construction and pitch contours in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 14. 225–236.

Uyeno, Tazuko, Hideo Hayashibe, Kunihiko Imai, Hiroshi Imagawa and Shigeru Kiritani. 1981. Syntactic structure and prosody in Japanese: A study on pitch contours and the pauses at phrase boundaries. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 15. 91–108.

Uyeno, Tazuko, Hiroshi Yamada, Hideo Hayashibe and Rieko Aoki. 1980. Comprehension of complement constructions in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 14. 237–262.

Venditti, Jennifer J. 2005. The J_ToBI model of Japanese intonation. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology and intonation of phrasing*, 172–200. Oxford/New York: Oxford University Press.

Venditti, Jennifer J., Kikuo Maekawa and Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 456–512. Oxford/New York: Oxford University Press.

Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: MIT Press.

V Broader perspectives

Tomoaki Takayama

15 Historical phonology

1 Introduction

Historical background is significant and useful for direct observation of language structure. This chapter provides information on historical issues helpful to understanding the synchronic aspects of the phonetics and phonology of modern Tokyo Japanese.

Due to a lack of established cognate relations with other languages or linguistic families, reconstruction of prehistoric stages of Japanese is quite often difficult. Thus, historical studies of the language are inevitably different from the studies of other languages, such as the Indo-European languages, which use the comparative method as a central tool. We have to depend largely on internal resources in the historical studies of Japanese. Fortunately, the historical stages of Japanese are attested back to about the eighth century, with a large amount of writing materials and dialectal information.

In order to shed light on what historical studies reveal about the structure of modern Japanese, this chapter addresses several historical issues and gives a review of the results and the points of controversy from studies over the last few decades. Note that it does not present a comprehensive review of the field nor a comprehensive outline from the earliest period to the present (see Frellesvig 2010 as well as the chapters in the History Volume of this handbook series for full discussion of the historical phonology of Japanese). It briefly mentions technical treatments of writing system and philological problems only to the extent these are helpful to understanding historical phonology.

Moreover, the focus of this chapter is on segmental aspects of phonology, not on suprasegmental or prosodic features such as tone (i.e., accent). Due to lack of sufficient historical data, the tonal system of modern (Tokyo) Japanese cannot be easily traced back to earlier stages, even though the tonal system of the Kyoto dialect can be attested in the twelfth century in written materials (see Kubozono 2012 and Uwano 2012 for the variety of pitch accent systems in the dialects). Intonation also complicates the situation. The tonal history of the compounds in modern Japanese is not yet well understood. For these reasons, prosodic aspects are not dealt with in this chapter, even though they are undoubtedly an integral part of the phonetic and phonological studies of Japanese.

This chapter is organized as follows. Section 2 discusses various issues relating to voiced obstruents, with main focus on the velar nasal [ŋ] and prenasalization in the consonantal system. Section 3 deals with the historical backgrounds of the affricates, [ts], [tʃ], [dz], and [dʒ], as well as the asymmetry these sounds exhibit in modern Japanese. In section 4, we discuss some problems concerning reconstruction

to point out that the phoneme /s/ of modern Japanese was realized as affricates in old Japanese. Section 5 briefly presents a history of the phonemes /h/ and /p/, based on recent research in this area. The final section (section 6) summarizes the main points of the chapter, and mentions residual issues for future studies.

In this chapter, the IPA symbol [u] is used to refer to the various phonetic realizations of the vowel /u/ of Japanese, and therefore does not exclude the unrounded realization that is represented by [ɯ] in regular usage. This broader representation is quite often adopted in the literature of historical phonology, since the accurate phonetic values cannot always be defined when discussing historical issues. The symbol [ɯ] is used in this chapter only when it is necessary to specify the unrounded realization.

2 Historical issues about voiced obstruents

When we discuss the Japanese consonant system and its history, we must define terminology. The English term "voiced obstruents" commonly used to refer to the phonemes such as /b/, /d/, /g/, and /z/ of modern Japanese is itself problematic, especially with respect to their phonetic realizations. In Japanese linguistics, what we call voiced obstruent is generally referred to as a *daku-on* consonant (see Kubozono's introduction to this volume). While the Japanese term is not well known except to specialists, this label is convenient for identifying a specific category in the consonant system without suggesting any kind of phonetic realization such as prenasalization. In modern Tokyo Japanese, *daku-on* consonants are largely pronounced as voiced obstruents, but this is not necessarily true of other dialects. Similar attention is required when comparing different historical stages of a dialect spoken in a specific area. For example, in the history of *daku-on* in the Kyoto dialect, it is necessary to refer to the category without phonetic specifications, as discussed in the following sections. In fact, one of the important issues about the consonants of this dialect is the de-prenasalization in the *daku-on* series: most members of this series pronounced as prenasalized consonants changed to the plain voiced obstruents. Even in modern Japanese, which is our main concern in this chapter, we face problems with phonetic realizations. The standard pronunciation of /g/ in non-initial position is the velar nasal [ŋ] (see section 2.1 for details). In these cases, the term "voiced obstruent" is not completely phonetically accurate, and should be avoided. However, since it is hard to find any other suitable English term, we will continue to use it, in a categorical sense, with no phonetic specification.[1] When clarification is required during discussion, we will add a modification, for example, such as 'modern voiced obstruents'.

1 Frellesvig (2010: 34–36) refers to the distinction between *sei-on* and *daku-on* in Old Japanese as tenuis (tense) versus media (lax), avoiding terms such as voiceless and voiced. Whether the tenuis were allophonically voiced in intervocalic position is a controversial question (see note 10 and Hayata 1977).

2.1 Variations of /g/

Modern Tokyo Japanese has the velar nasal [ŋ] among the variations of the phoneme /g/, as mentioned above. From the prescriptive standpoint, this nasal sound has been regarded as one of the significant elements that characterize the standard pronunciation of Japanese for a long time. The instructions to carefully keep on pronouncing the nasal for the non-initial /g/ have been repeatedly given and emphasized especially in broadcasting, singing, and classes in schools, although nowadays such prescriptive manifestations are not observed quite as often as before (see Vance 1987: 108–132; Vance 2008: 214–222; and Shibatani 1990: 171–173 for details). In this section, we will focus on the synchronic and diachronic aspects of the allophonic nasal sound.[2]

To begin with, we show the variations of /g/ in modern Tokyo Japanese. The initial allophone is exemplified in (1). For non-initial positions, there are two cases: the intervocalic position as shown in (2) and (3), and the postnasal position as shown in (4).

(1) gomi [gomi] 'trash'

(2) tamago [tamaŋo] 'egg'

(3) kagi [kaŋi] 'key, lock'

(4) ringo [riŋŋo] 'apple'

The oral stop [g] occurs in the initial position, whereas the nasal [ŋ] appears in non-initial positions. The oral and the nasal occur complementarily according to the position. As opposed to /g/, other voiced obstruent phonemes shown in Table 1 have no nasal variants. That is, /b/ is not phonetically realized as [m], and the phonemes /d/ and /z/ are not realized as [n]. Among the four voiced obstruents in Table 1, /g/ is unique in having the nasal realization.

Table 1: Voiced obstruents and nasals in modern Japanese

	labial	dental	velar
voiced obstruents	b	d	g
		z	
nasals (nasal onsets)	m	n	

(other phonemes omitted)

2 There is a well-known controversy in the literature about whether the difference between [g] and [ŋ] should be treated as allophonic or phonemic. In this chapter, we will not elaborate on this problem nor on the morphophonological issues involved (see Vance 1987: 108–132; Komatsu 1981: 137–148; among others for surveys and comments).

The difference in the behavior of /g/ and other voiced obstruents can be easily explained by the absence of the nasal /ŋ/ in the onset. In the labial and dental places, the nasal onset /m/ and /n/ exist, respectively. If each phoneme shares the same place of articulation with its nasal counterpart, it does not allow for any nasal variant. Since there is no gap in any other place than the velar, there is no room for /b/, /d/, and /z/ to be realized as nasal variants. On the other hand, /g/ permits the variant [ŋ] to fill the accidental gap in the nasal place: that is, the gap allows nasal realization of the phoneme /g/.

However, the allophonic range between the oral sound and the nasal is not straightforwardly understandable in terms of phonetic contextual effects. We cannot find any reason why nasality should be introduced to the realization of the non-initial /g/ (Hattori [1957] 1960: 338–341; Vance 1987: 111–112). Of course, in the case of the postnasal position such as *ringo* in (4), the nasal variant is likely to occur due to assimilation to the preceding nasal consonant. In the intervocalic environment, by contrast, a variant need not be a nasal sound. It is hard to understand why the nasal allophone [ŋ] occurs synchronically. Instead, a plausible allophone is the fricative [ɣ] because the intervocalic position often causes spirantization. Actually, we quite often observe this fricative rather than the nasal [ŋ] in the present pronunciation of Tokyo Japanese. For example, *tamago* in (2) is realized as [tamaɣo], and *kagi* in (3) is as [kaɣi] (Vance 1987: 111–112; Kindaichi 1942; among others). In order to understand the intervocalic nasal realization in modern Tokyo Japanese, it is necessary to look at the diachronic background.

2.2 Diachronic background of velar nasal

Due to scarcity of resources, it is not feasible to trace the history of the Tokyo dialect back more than several hundred years. As for the Kyoto dialect, it is readily known that the modern voiced obstruents were realized as prenasalized consonants in the beginning of the seventeenth century, as will be discussed in section 2.3. A question arises about whether Kyoto is a special case or if similar situations are observed generally in Japanese.

When we look at the realization of /g/ in the consonant systems of the dialects, we quite often encounter nasal variants. While a number of dialects have no nasal but oral [g] or [ɣ], sounds such as [ŋ] and [^{n}g] are widely observed throughout Japan. Thus, the prenasalization in the Kyoto dialect of the seventeenth century is not idiosyncratic. Inoue (1971) discusses issues of the phoneme /g/ in various dialects, analyzing the nationwide geographical distribution and providing a diachronic perspective. Synthesizing the geographical information and the historical facts demonstrated by the document resources, Inoue (1971) concludes that the great majority of dialects can be interpreted by two events and their relative chronological order: (i) the emergence of the variant [ŋ], and (ii) the de-prenasalization, i.e., the vanishing

of the nasal element of prenasalized consonants. If [ŋ] emerged in advance, that variant is expected to remain even if the de-prenasalization were completed in the following stage. By contrast, if the de-prenasalization took place in advance, i.e., if the prenasalized [ⁿg] changed to the oral [g] or [ɣ]; the nasal [ŋ] cannot emerge in any following stage, except for the case of contact with another dialect. To sum up, the allophonic [ŋ] may be derived from the prenasalized realization of /g/, in a way that it fills the velar nasal gap in the consonant system, as we saw in Table 1 (see Kamei 1956 for historical discussions on the variation of /g/). Consequently, the nasal [ŋ] in Tokyo Japanese can be regarded as a relic of the phonetic quality in the earlier stage of the consonant system.

2.3 Attestation of prenasalized consonants

The prenasalization in the Kyoto dialect is attested in historical resources. The most significant document is a handbook of the Japanese language for foreign learners, *Arte da lingoa de Iapam*, written by João Rodriguez, and published in Nagasaki by the Society of Jesus in 1604–1608. According to his instruction on the Romanization of Japanese (Rodriguez 1604–1608: 177–178 and Doi 1955), the vowel letters such as *i*, *e*, *a*, *o*, and *u* followed by the letters *d* and *g* should be pronounced with the half size of the *til*(tilde) '˜', which is considered to suggest prenasalization (Hashimoto 1932). His instruction also says that the learners must not pronounce them with the distinct 'til' (Rodriguez 1604–1608: 172). For example, the spelling *toga* 'offense, sin' should not be pronounced as *tonga* equivalent to *tõga* including the complete tilde that indicates the coda nasal. Since the relationship between the orthography and the sounds is complicated, we summarize the correspondences in (5).[3]

(5) a. The letter *d* is used in *de*, *da*, *do*, and *dzu* which represent the Japanese sounds [de], [da], [do], and [dzu], respectively. Note that the absence of [du] and [di] is due to the affrication (see section 3.1). For [ʤi], it is represented by the spelling *gi* in (5b).

b. The letter *g* is used in *gui, gue, ga*, *go, gu*, and *gi* which represent the Japanese sounds [gi], [ge], [ga], [go], [gu], and [ʤi], respectively.

According to (5), the prenasalized realization occurs in the stops and the affricates such as [d], [g], [dz], and [ʤ]. As for the letter *b* that corresponds to the phoneme /b/, Rodriguez remarks that the half size of the tilde is observed in some cases, but it is not so common compared to the extent of the letter *d* and *g*. His comment suggests that the prenasalization in the labial was weaker than in the dental and the velar. Furthermore, in the revised concise version *Arte breve da lingoa de Japoa*

3 The romanization of Japanese in works by the Society of Jesus is based on the orthography of Portuguese.

published in 1620, Rodriguez mentions that the half size of the tilde was occasionally observed before the letters *j* and *z* which correspond to the phoneme /z/ (Rodriguez 1620: 12). By contrast, there are some domestic resources suggesting that /z/ lacked prenasalized realization (see section 3.3).[4]

Rodriguez also refers to the fact that the letter *g* in the Bizen dialect lacked the half size of the tilde and was pronounced *secamente* 'with a dried sound' (Rodriguez 1604–1608: 171). The sound is considered to be the non-nasalized realization [g] or [ɣ] of /g/. Bizen is a part of what is now Okayama Prefecture, situated in the outer area adjacent to Kinai, the central region where the capital city Kyoto was centrally located. His comment on the dialect reveals an interesting sociolinguistic aspect concerning the velar sound. It presumably suggests that the habitants of the capital were sensitive to the rural accent and that the lack of prenasalization made a harsh auditory impression on them.

A subsequent stage of the Kyoto dialect is demonstrated by domestic documents written about one hundred years later after Rodriguez (1604–1608). The instructions in these documents advise the readers to keep the prenasalization in /di/ and /du/ (see section 3.3). It reveals that de-prenasalization had already taken place. Moreover, in those texts, the authors applied the terms originally referring to the coda nasal /N/ to the nasal element of prenasalization. A representative of these documents is *Ikeisai Kōgo Kikigaki*, which is assumed to date back to the beginning of the eighteenth century. It instructs that the word *midu* (/du/ was realized as [dzu], see sections 3.1 and 3.3 for the phonetic realization), which is the earlier form of *mizu* 'water' in Modern Japanese (ModJ), should be pronounced with a shorter coda nasal /N/ inserted before /du/. This instruction in the manuscript *Ikeisai Kōgo Kikigaki* resorted to the Kana script 'ん', which normally represents the coda nasal, in order to refer to the nasal element that should be inserted: "*midu* 'water' should be pronounced in a similar way to 'みんづ' <mi-N-du>".[5] It is natural that the elder generation who originally maintained the prenasalization perceived it not as two successive sounds but as an inseparable sound [^{n}d]. Stated conversely, the younger generation who had not acquired the prenasalization perceived it as the sequence of a coda nasal and a plain voiced obstruent instead of an inseparable consonant. The notation of *Ikeisai* suggests the recognition by younger generations (T. Takayama 1998).

2.4 Velar nasal in the past

Although the status of the velar nasal as standard has not yet been lost, its prescriptive restraint is not imposed nowadays as strictly as before.[6] Based on his investiga-

---

4 Yamane-Tanaka (2005) discusses significant aspects of the de-prenasalization in the framework of Optimality Theory.

5 The purpose of this instruction is to preserve the traditional recitation style of *waka* poems.

6 On the basis of his observations, Vance (1987: 111) pointed out that the difference between the prestige norm of [g] and the official status [ŋ] can explain the preference of native speakers.

tion, Kindaichi (1942) pointed out that the younger generation was going to lose the velar nasal, and predicted that the sound would vanish in the future, no matter how much the non-nasal [g] was corrected in language education.[7] In fact, a half century later, the population maintaining the velar nasal allophone decreased by half in the past several decades (Hibiya 1988, 1995, 2002; Inoue 1983, 1998: 162–167; among others).

Transition from the velar nasal to the oral stop may be inevitable in the consonantal system of Japanese, but, at the same time, we cannot discount the fact that the nasal variant has lasted a long time. This fact must be discussed from a morphophonological (see note 2; Komatsu 1981: 145–148; Ito 1997) or sociolinguistic perspective.

As for the sociolinguistic viewpoint, fortunately, we have a historical resource that tells us about an older situation in Tokyo (Edo) Japanese. The comic tale *Ukiyo buro*, which was published in the beginning of the nineteenth century, depicts vivid conversations among ordinary people enjoying a public bath. In one of the scenes, an attendant of the bathhouse speaks in his rural dialect, contrastively different from the urban speech. The author *Shikitei Sanba* not only adopted the rustic vocabulary but utilized some speech sounds in order to emphasize the rural character. For that purpose, he marked the kana scripts representing the syllables with the velar /g/ with a special diacritic, i.e., a small circle instead of the usual two dots *dakuten* that indicates /g/. It is assumed that the special circle represents the non-nasal [g] or [ɣ] that might cause a harsh sensation or an unsophisticated impression to the urban native speakers.[8] This fact should be taken into consideration when we discuss the background of the status of the standard nasal [ŋ] in modern Japanese.

2.5 Prehistory of voiced obstruents

The works by Joan Rodriguez provide direct evidence for prenasalization in the beginning of the seventeenth century, as mentioned in section 2.3. However, it is difficult to find documents directly indicating the phonetic value in earlier stages. Nevertheless, on the basis of indirect resources and lack of contradicting evidence, it is generally assumed that voiced obstruents were largely realized with prenasalization even in Old Japanese (M. Takayama 1992, 2012: Ch. 3).

In prehistoric stages, we can do nothing but depend on hypothetical approaches, since there is no concrete evidence available to the reconstruction of the phonetic values. As far as the phonotactics is concerned, there is a noticeable fact that they occur in the native lexicon under the two distributional restrictions shown in (6).

---

7 Shibatani (1990: 171) remarks on the significance of Kindaichi's (1942) contribution to sociolinguistic studies.

8 Sakanashi (1975) claims that the special diacritic does not represent the oral sounds of /g/.

(6) a. Voiced obstruents do not occur in word-initial position.
b. There is a maximum of one voiced obstruent per morpheme.

These phonotactic restrictions in modern Japanese are well known in the literature, and they operated even in the eighth century, the earliest period to which we can date back by documental resources (see Ito and Mester 1986 for a theoretical discussion; Kamei 1970a; Morita 1977; Yamaguchi 1988 for historical discussions). Therefore, these distributional facts should be taken into consideration even in the hypothetical approaches. A plausible scenario is that such phonotactic restrictions come from some diachronic processes. In addition, the genesis of these distributions is considered to relate to the history of the voiced obstruents, i.e., prenasalized consonants.

Another important fact concerning these restrictions is that they are closely related to the mechanism of sequential voicing known as *rendaku* (see Vance, this volume, for full discussion). In other words, the phonotactic restrictions in (6) not only govern each simplex word in the native lexicon but also play significant roles at the morphophonological level. Actually, rendaku formation quite often took place even in the eighth century. We briefly show the relationship between rendaku formation and phonotactic restrictions, with modern compounds in (7) and (8). The same mechanism is true of the attested earliest stages.

(7)	a.	ude	b.	tamesi	c.	ude-damesi
		arm		try.GER		arm-try.GER
		'arm'		'trial'		'trial of skills'

(8)	a.	ude	b.	kurabe	c.	ude-kurabe
		arm		compete.GER		arm-compete.GER
		'arm'		'competition'		'competition of skills'

According to Komatsu (1981: 101–107), rendaku formation is established on the basis of the restriction in (6a). Namely, the non-initiality provides the following two merits in rendaku formation. First, the simplex *tamesi* in (7), which is subject to the restriction in (6a), changes to the form *damesi* when placed in non-initial position in the compound in (7c). The alternation to the voiced obstruent /d/ shows that the morpheme {tamesi} in the compound in (7c) is not in word initial position any longer. Note that the scope of (6a) is the word, not the morpheme, and therefore, (6a) is valid not only for simplex but also for compound words. Second, there is no minimal pair between a voiceless obstruent and its voiced counterpart in the initial position due to the restriction in (6a). Therefore, this gap makes easy recovery to an original simplex from the form initially voiced by rendaku.

The phonotactic restriction in (6b) blocks rendaku voicing. This process is well known as Lyman's Law or Motoori-Lyman's Law: for details, see Vance (this volume); Vance (1987: Ch. 10); Yamaguchi (1988); Ito and Mester (2003); van de Weijer, Nanjo, and Nishihara (2005); among others. If a morpheme has a voiced obstruent in medial position, rendaku voicing does not occur in the morpheme, as shown in (8). The simplex *kurabe* in (8b) has one voiced obstruent, and therefore cannot undergo rendaku, even if this form follows the lexical element *ude* in (8a), as in the compound *ude-kurabe* in (8c). By contrast, since the simplex *tamesi* has no voiced obstruent, as shown in (7), it can and actually does undergo rendaku voicing, as in the compound *ude-damesi* in (7c).

In sum, the distributional properties of voiced obstruents have a close relationship with the morphophonological aspect, which should date back to the prehistoric period. Hizume (2003) proposes a scenario in which the Japanese consonantal system had only one series of obstruents at some earlier prehistoric stage that later bifurcated. According to this view, the prenasalized consonants arose in the initial positions of the second lexemes in compounds in order to denote concatenation and demarcation in compounds; and, thus, the first stage of rendaku emerged.[9] On the other hand, the plain obstruents were phonetically voiced in word-medial position, and their voiced realization was different from prenasalization from a functional viewpoint, while the voiced realization later became weakened in the dialect of Kinai, the central area (cf. Hayata 1977).[10] In this way, when we discuss the history of the voiced obstruents in Japanese, the focal point is how we deal with rendaku and the distributional properties in (6).

3 Affrication and merger

3.1 Affrication

When we look at the dental stops in modern Japanese, [t] and [d] do not occur before the high vowels /i/ and /u/ in native and Sino-Japanese (henceforth SJ) words; in other words, there are no syllables such as [ti], [di], [tu], and [du]. In the positions preceding these two vowels, affricates occur instead. These distributional gaps in the dental stops result from a historical change that took place in approximately the sixteenth century. Concretely, before the front vowel /i/ and the palatal

9 Hizume (2003) does not completely attribute the prenasalized consonant occurrences in the native lexicon to the genesis of rendaku. He proposes the historical stratification of the prenasalized consonants (i.e., voiced obstruents) in the native lexicon. According to this proposal, there are prenasalized consonants that emerged after the bifurcation, in addition to their predecessors.

10 Hayata (1977) argues that the voiceless obstruents (in a categorical sense) of the Kinai dialect were phonetically voiced in intervocalic positions even in the Heian period, probably around the eleventh century (see also M. Takayama 1992, 2012: Ch. 3; Frellesvig 2010: 34–36).

glide /j/, the stops [t] and [d] changed to [tʃ] and [dʒ], respectively. In addition, before the back vowel /u/, [t] and [d] changed to [ts] and [dz], respectively. As a result, in modern Japanese, /ti/ is phonetically realized as [tʃi], and /tu/ as [tsu] (~[tsɯ]). Kim (2001) and Lin (2011) mention these Japanese affricates in their cross-linguistic discussions, but they do not consider the backgrounds particular to Japanese. We discuss some details of the affrication with focus on the historical and structural contexts in which this sound change took place.

The affrication in question is noteworthy with respect to the environments under which it took place. If we treat this change uniquely, we face difficulty in terms of the phonetic motivation. Generally speaking, the frication before the front vowel (and the glide /j/) is likely triggered by palatalization due to assimilation to the following vowel /i/. However, this explanation cannot apply to the affrication before /u/. Accordingly, we would need to deal separately with these two affrication paths in phonetic discussions.

As for the unrounded realization of the vowel /u/ (see Kubozono's introduction to this volume), we notice that /u/ shows a notable tendency to centralization when following non-palatal sibilants such as [s], [z], and [ts] in modern Japanese. (Note that the difference between the voiced fricative [z] and the affricate [dz] is not distinctive in modern Japanese: see section 3.2 for historical background). /u/ in those contexts is often described as [ɯ̈] in IPA, such as [sɯ̈], [zɯ̈] and [tsɯ̈]. If the centralization can also trigger frication, this may explain the affrication before /u/.

At the same time, however, it is necessary to figure out why each phonetic condition simultaneously caused different types of affrication in the history of a single language. If the two events did not take place accidentally in the same period, we must consider other aspects in order to understand the events as one and the same process. Although the affrications of the dental stops ([t] and [d]) had different phonetic motivations, it is not realistic to deny the uniformity of a historical event.

In modern Japanese, the difference between the two affricates [tʃ] and [ts] is not distinctive, at least in the native vocabulary, since their choice always depends on the following vowel: /i/ or /u/ (see Kubozono Ch. 8, this volume, for exceptions in loanwords and Pintér, this volume, for the introduction of new sound sequences in modern Japanese). As opposed to [ts], the palatalized realization [tʃ] is due to regressive assimilation to the front /i/ and the glide /j/ (Hattori [1955] 1960: 288, [1956] 1960: 321–322).[11] The difference between [tʃ] and [ts] is regarded as phonemically redundant. However, the practical role that the consonantal difference fulfills in

11 Hattori (1955, 1956) describes the affricate [tsu] as /cu/, and [tʃi], as /ci/; namely, he regards both sounds as derived from the phoneme /c/ on the grounds that the latter is palatalized due to the assimilation of the following front vowel /i/ and the glide /j/. On the other hand, he argues that since the difference between the stop [t] and those affricates cannot be straightforwardly explained by the phonetic environmental conditions, the two phonemes /t/ and /c/ are required in the description of the modern consonant system.

distinguishing between the syllables /ti/ and /tu/ should not be underestimated. The qualitative difference between these consonants provides an effective phonetic cue to the recognition of the resultant syllable. In addition, the high vowels are quite often dropped, due to high vowel devoicing (see Fujimoto, this volume, for details). Vowel deletion (or devoicing) frequently occurs in /ti/ and /tu/, too. For example, *tikai* 'oath' is usually realized as [tʃkai] or [tʃi̥kai], and *tuta* 'ivy', as [tsta] or [tsɯ̥ta]. In such devoiced realizations, the fricative part, such as [ʃ] and [s], is an indispensable element for recognizing which syllable is intended, /ti/ or /tu/, since there is no vowel that can carry out the distinctive function. These fricative parts, rather than the vowels, play the essential discriminating roles. Even in words where the vowel /i/ or /u/ is not completely dropped, the qualities of the fricative parts provide an important cue to the distinction between /ti/ and /tu/ (see section 3.2 for voiced obstruents). According to the phonemic interpretation (Hattori 1955, 1956), the consonantal difference between the palatal sound [tʃ] and the non-palatal [ts] is redundant, since it is automatically determined by the vowel that follows it. However, the contribution of fricative parts to the distinction should be regarded as vital.

A similar situation is observed in the differences between the palatal and dental fricatives, such as between [ʃ] and [s] as well as between [ʒ] and [z]. The two sounds of each set are allophones in the phonemic treatment, since the palatals, [ʃ] and [ʒ] always accompany the front vowel (and the glide /j/) and the non-palatal [s] and [z] do not occur before the front vowel. Nevertheless, the difference between palatal and non-palatal consonants is significant, notably when the vowels are dropped or devoiced. For example, *sita* 'tongue' is usually realized as [ʃta] (or [ʃi̥ta]); and *sukiyaki*, as [skijaki] (or [sɯ̥kijaki]).

The function that the fricative parts carry out in modern Japanese is noteworthy also in the discussion on historical affrication. Of course, it is difficult to demonstrate vowel devoicing or weakening by historical resources, but we refer to the comments on the pronunciations of Japanese in *Ars Grammaticae Iaponicae Linguae* of Didaco Collado, published in 1632 (see Ōtsuka 1957). It says that the vowels *i* and *u* in word final position is hardly audible for beginners; for example, the word "gozàru" (*gozaru* 'stay.HON') sounds like "gozàr", the word "fitòtçu" (ModJ. *hitotu* 'one') like "fitòtç", and the words "àxi no fàra" (ModJ. *asi no hara* 'field of reeds') like "àx no fàra". The latter two illustrations, in which the final /u/ or /i/ is deleted, are particularly interesting for our discussion here. Although second language resources need to be dealt with carefully, they may indicate devoicing or weakening of vowels in Japanese in the seventeenth century.

The emergence of the fricativisations in /ti/ and /tu/ (as well as /di/ and /du/) should be investigated taking into consideration phonetic cues, as mentioned above. According to T. Takayama (2006, 2009), this change is recognized as a historical trend toward activating the potential contrast in quality between palatal and non-palatal consonants. Without considering such phonetic differences, we cannot treat the twofold phenomenon that consists of the two processes, [ti]>[tʃi] and [tu]>[tsu]

(as well as [di]>[dʒi] and [du]>[dzu] in the voiced counterparts) as a unique historical event. The status or role played by the phonetic difference between palatal and non-palatal consonants should be further discussed from both synchronic and diachronic viewpoints in the future.

The affrication in question involves a chronological problem, too. As mentioned above, the two processes of affrication took place simultaneously, or more precisely, at almost the same time. Looking into the details, the affrication in /ti/ and /di/ is presumed to have slightly preceded the affrication in /tu/ and /du/. Historically, following these two affrications, two pair mergers took place in the voiced obstruents. Namely, the two oppositions, /di/:/zi/ and /du/:/zu/, disappeared at the next stage (see section 3.2). There is evidence to support the claim that the merger of /di/ and /zi/ took place a little earlier than that of /du/ and /zu/. We find three types of systems among the various dialects with regard to these two oppositions: (i) no oppositions are maintained, as seen in modern Tokyo Japanese; (ii) both oppositions are maintained, as observed in the Kōchi dialect; and (iii) only /du/:/zu/ is maintained, as in the Ōita dialect. In contrast, we do not find a fourth type where only /di/:/zi/ is maintained (Itoi 1962; Kuno et al. 1995; Kuno 2006; Sugimura 2001; among others). In addition, in the dialects which still preserve the opposition(s), it is often observed that the affrication in /tu/ and /du/ is not complete, compared to that in /ti/ and /di/. If it is true that there was a time lag between the two affrications, we suggest that the affrication triggered by palatalization promoted the other affrication (T. Takayama 2009). Such a time lag is remarkable even considering general tendencies about affrication. Further cross-linguistic investigations into similar cases are needed.

As for the chronological problem, another question arises: why did this change take place around the sixteenth century? This question has a close relation to the issues regarding the phonetic development of /s/. It is generally assumed in the literature (see section 4) that the phoneme /s/, and probably /z/, were realized as affricates, at least in part, in earlier stages. The transition in /s/ and /z/ from the affricate to the fricative must have taken place at some chronological point before /t/ and /d/ were affricated before /i/ and /u/. Ogura (1998) discussed the relevant chronological issues both from the structural and diachronic viewpoints.

3.2 Merger

There is an asymmetry in the sibilants between the voiceless and voiced series. While the difference between the voiceless fricative and the voiceless affricate is contrastive, as illustrated by the minimal pairs in (9) and (10), the voiced obstruents lack the contrast between the fricative and the affricate. This difference is summarized in Table 2: Parenthesized palatal sounds such as [ʃ], [tʃ], [ʒ], and [dʒ] occur before the front vowel /i/ or the glide /j/.

(9) a. siru ([ʃ]) : tiru ([tʃ])
'learn, know' 'scatter, fall'

b. husi : huti
'node, joint' 'edge, brim'

(10) a. sumi ([s]) : tumi ([ts])
'corner' 'guilt, sin'

b. hanasu : hanatu
'speak, talk' 'shoot (an arrow or a bullet)'

Table 2: Asymmetry between voiceless and voiced obstruents in modern Japanese

	fricative	affricate	difference between fricative and affricate
voiceless obstruent	[s] ([ʃ])	[ts] ([tʃ])	contrastive
voiced obstruent	[z]([ʒ])	[dz]([dʒ])	non-contrastive

The asymmetry results from the mergers, as discussed in section 3.1. As far as the Kyoto dialect is concerned, the contrast between the two kinds of voiced obstruents was confused to a large extent at the end of the sixteenth century, and completely merged in the next century. Namely, /di/ ([dʒi]) merged with /zi/ ([ʒi]), and /du/ ([dzu]) merged with /zu/ ([zu]). For example, the word *kuzu* 'trash' in modern Japanese comes from /kudu/ of the earlier stage, spelled 'くづ' in kana script, and the word *kuzu* 'kudzu vine' comes from /kuzu/, spelled 'くず' in kana script (the English spelling of "kudzu" bears no relation to the original kana spelling).[12] While these words are homonyms in modern Japanese, they were distinguished from each other in the earlier stage. On the other hand, the voiceless contrasts /ti/:/si/ and /tu/:/su/ are still preserved, as shown in (9) and (10).

Considering the chronological fact that the affrication took place in the period just prior to the merger, the merger was no doubt caused by the affrication. The affrication before the high vowels made the qualitative distance between /d/ and /z/ closer, and eventually the opposition became neutralized in those positions. Note that the opposition between /d/ and /z/ itself is maintained, since the affrication did not occur with other vowels such as /e/, /a/ and /o/; namely, the contrasts /de/:/ze/, /da/:/za/, and /do/:/zo/ are preserved. As to the partial merger of /d/ and /z/, there is a historical question that should be explained: why is the consequence in the voiceless series different from that in the voiced series? The affrication took place in the voiceless [t] as well as in the voiced [d] and, therefore, /t/:/s/

12 Although pre-modern kana letters were quite often spelled without the diacritical mark *dakuten* indicating a voiced obstruent, we will show here the kana spelling with *dakuten* added.

should have been subject to the same phonetic condition that the qualitative distance had become closer; nevertheless, the voiceless opposition /t/:/s/ has been entirely preserved in modern Japanese. The asymmetry pointed out above concerns not only the system of modern Japanese but also the process of the historical change. This is an interesting theoretical point for phonetics and phonology in general, as well as for Japanese in particular, some relevant issues of which are discussed here.

First, while there are fricative variations among the realizations of voiced obstruent phonemes, there is no voiced phoneme that is exclusively realized as fricative in modern Japanese. For example, the phoneme /b/ is realized not only as the stop [b] also as the bilabial fricative [β] especially in intervocalic positions. Moreover, the phoneme /z/ is realized not only as the fricative [z] but as the affricate [dz]. In contrast, some voiceless phonemes are exclusively realized as fricatives, such as /s/ and /h/.[13] When we look at the voiceless consonants in the stage around the fifteenth century, the period just before the affrication and merger, it is probable that there were two fricative phonemes, /s/ and /ɸ/, in the system. Among the voiced (or prenasalized) obstruents, on the other hand, there is only the phoneme /z/ as a candidate that would have been exclusively realized as fricative; however, since there is no clear evidence, it is difficult to establish such reconstruction. In addition, it is assumed in the literature that /z/ was realized mainly as prenasalized consonants (probably affricates) in old Japanese (see sections 3.3 and 4). If the situation had not changed in the fifteenth century, there would be no phoneme realized exclusively as voiced fricative in the consonant inventory. Thus, the voiced obstruents compared to the voiceless obstruents do not have fricative sounds that bear a distinctive function in the system. This fact should be taken into consideration when discussing the asymmetry in the merger, which may come from the asymmetry in the consonant system.

A second issue concerns the difference in the functional load between the voiceless contrasts and the voiced contrasts, shown in (11).

(11) a. voiceless contrasts: /ti/:/si/ (including /tj/:/sj/), and /tu/:/su/
b. voiced contrasts : /di/:/zi/ (including /dj/:/zj/, and /du/:/zu/

The distribution of voiced obstruents is constrained by two morpho-phonotactic restrictions in the native lexicon, as discussed in (6) above. These restrictions have lasted through the known history of Japanese. Thus, the voiced obstruents occurred less frequently than the voiceless obstruents, which suggests that the functional load of the voiced contrasts was fairly low. In fact, there were only a few minimal pairs for the voiced obstruents in (11b) in the lexicon of Japanese of the sixteenth century

13 Whether /h/ is a typical fricative or not is problematic, but at least, it is not definitely realized as a stop.

(T. Takayama 1993).[14] A contrast may be vulnerable when the functional load is extremely low. The functional load may not always be a decisive factor in historical events, but we must seriously consider it when discussing the merger in the series of voiced obstruents.

3.3 Merger and prenasalization

Another topic about the merger quite often discussed in the literature concerns the historical relation to the loss of prenasalization. Against the usual confusion caused by merger, historical documents give instructions about the prescriptive pronunciation as well as how to correct deviations there from. Such documents often appeared from the end of the seventeenth century to the beginning of the eighteenth century. They instructed that the distinctions between /di/ /du/ and /zi//zu/ should be made in the way shown in (12).[15]

(12) a. /di/ and /du/ should be pronounced with a shorter coda nasal immediately before them, and with the tip of tongue touching to the roof of the mouth.

b. /zi/ and /zu/ should be pronounced with no shorter coda nasal, and with the tip of tongue not touching to the roof of the mouth.

The instructions in (12) reveal the earlier stage just before the distinction was lost. The phonetic difference is described in (13).

(13) a. /di/ and /du/ were realized as prenasalized affricates.
b. /zi/ and /zu/ were realized as plain voiced fricatives.

The instructions suggest, first of all, that the prenasalization was being lost in the seventeenth century (see section 2.3 for detailed discussion), and secondly, that the prenasalization of /z/ vanished earlier than that of /d/ and /g/. The time lag eventually provided a chance for the distinction between /d/ and /z/ to be carried by the difference in prenasalization, in addition to the difference between affricates

14 There were two minimal pairs of words sharing the same pitch accent of the Kyoto dialect in the sixteenth century. One is a pair of native words, *udi* 'family, clan' and *uji* 'maggot', and the other is a pair of the initial positions of the SJ words, *di* 'ground' and *ji* 'letter, Chinese character'. The words mentioned in the text, *kudu* 'trash' and *kuzu* 'kudzu vine', do not have the same pitch accent.
15 Although there are some differences among the kinds of instructions, we will not give the details and differences that are observed among documents. The instruction shown in (12) is a summarized version (see T. Takayama 2003 for details of differences).

and fricatives. In other words, the difference of prenasalization reinforced the distinction between affricates ([dz][dʒ]) and fricatives ([z][ʒ]). This is summarized in Table 3, which schematically illustrates phonetic realizations in intervocalic positions. The variants, [dʒi] and [dzu], occurred in the initial positions and after the coda nasal /N/ in the fourth stage.

Table 3: The processes of the mergers between /di//du/ and /zi//zu/

	/di//du/	/zi//zu/
1. the initial state	[ⁿdi][ⁿdu]	[ⁿdʒi][ⁿdzu]
2. the stage after the vanishing of the nasal portion in /z/	[ⁿdi][ⁿdu]	[ʒi][zu]
3. the stage after the affrication of /d/	[ⁿdʒi][ⁿdzu]	[ʒi][zu]
4. the stage after the merger resulting from the loss of the nasal portion in /d/	[ʒi][zu]	[ʒi][zu]

As a result, the contrast between /di/:/zi/ and between /du/:/zu/ temporarily resisted merger. In this way, the de-prenasalization and the merger accidentally overlapped in the historical context. A kind of accidental synchronization between independent diachronic events reveals dynamic characteristics of the phonological history. See Kamei (1950), M. Takayama (2006, 2012: 147–162) and T. Takayama (1993, 2010) for historical discussions of the processes shown in Table 3; see Steriade (1993) and Riehl and Cohn (2011) for theoretical discussions on the relationship between prenasalization and the affricates.

4 Issues concerning the phonetic realization of /s/

The phonemic status of /s/ has been relatively stable with no drastic changes until modern Japanese; that is, we do not find any merger or split through history, apart from some speculations mentioned about the prehistoric period (section 4.1). In modern Japanese, the phoneme /s/ is realized exclusively as fricatives [s] and [ʃ]. However, investigation into the phonetic value of /s/, dating back to the earliest stage attested by documented resources, shows a different situation. It is widely accepted that the phonetic realizations of /s/ in the eighth century can be reconstructed as affricates, or at least, basically as affricates (Arisaka 1936; Kamei 1970b; Mori 1991; Ogura 1998; Takeuchi 1995; Hayashi 2002).

Since the affrication of /ti/ and /tu/ took place in the sixteenth century, as discussed in section 3 above, it is assumed that the transition of the phonetic realizations of /s/, from affricates to fricatives, should have taken place in some earlier time before the sixteenth century, although attestation of an accurate date is extremely difficult.

In this section, we note a few problems that arise as consequences of old Japanese reconstruction of /s/. Although technical discussions about historical data analysis or philological issues are relevant to such reconstruction, they are not discussed here (see the History Volume). The focus here is on the phonological aspects of the reconstruction of /s/.

4.1 /s/ in the consonant system of Old Japanese

The first problem concerns how the phonemes are organized in the consonant system of Old Japanese. Since it is generally assumed in the literature that the phonemes that correspond to /s/ and /z/ in modern Japanese were realized as affricates, we will use the symbols /ts/ and /dz/ to represent them. The consonant system of Old Japanese is shown in Table 4.

Table 4: The consonant system of Old Japanese

voiceless (or non-prenasalized)	p	t ts	k
prenasalized	b	d dz	g
nasal	m	n	
liquid		r	
approximant		j	w

We briefly comment on each member of the consonant system in this table. First, the labial /p/ corresponds to /h/ of modern Japanese. The fact that the modern /h/ comes from the labial sound is unquestionable in the literature (section 5). Second, /d/, /g/, and /b/ at the beginning of the seventeenth century had prenasalized realizations, as discussed in section 2.3. Whether Old Japanese had prenasalized /b/, /d/, /g/, and /dz/ is not well-documented, but there is no negative evidence against such realizations (sections 2.3 and 2.5). Therefore, these segments are generally recognized as prenasalized in the earliest attested stage. For the other phonemes /m/, /n/, /r/, /j/, and /w/, there is essentially no controversy about their phonemic status and phonetic values, and they are thought to be essentially no different from their modern Japanese counterparts.

A look at the obstruents in the inventory shows a remarkable characteristic: there is no fricative but there are affricates such as /ts/ and /dz/. For the voiceless obstruents, except for /ts/, there are the stops /p/, /t/, and /k/, as discussed at the beginning of section 4. The situation of the voiced obstruents (or the prenasalized obstruents) may be the same as that of the voiceless ones. Note that the prenasalized realization may associate closely with the realizations of the stops (see Steriade 1993 and Riehl and Cohn 2011).

Regarding the voiceless obstruents, Kamei (1970b) pointed out that the reconstructed system does not accord with a cross-linguistic implication about the affricate: if there is an affricate in a consonant system, then there should be a fricative as its counterpart. Namely, if the affricate /ts/ existed in the system, the fricative /s/ would be expected to exist, too. However, the attestations from historical resources do not indicate that the fricative /s/ existed in Old Japanese, apart from /ts/, which was a predecessor of modern /s/. Arisaka (1955: 489–490) speculates that the fricative /s/ might have existed in a prehistoric stage, but later vanished. Possibly, the initial /s/ vanished via /h/, and the intervocalic /s/ merged into /ts/, which would often be weakened in intervocalic positions. Kamei (1973a) discusses a morphological phenomenon that may support such a hypothesis, that is, the alternation between the zero consonant and the /ts/ as observed between the simplex /ame/ 'rain' and the derived form /tsame/ in compounds such as *haru-same* 'spring rain' in modern Japanese.[16]

As Kamei (1970b, 1973a) points out, the gap of the dental voiceless fricative in the consonant system of Old Japanese remains a mystery in the historical studies of Japanese phonology. Cross-linguistic approaches may help us solve the mystery. State-of-the-art theoretical studies may impact on the reconstruction of the affricate in Old Japanese.

As for the details of phonetic realizations, it should be mentioned that the proposed reconstruction does not necessarily exclude the possibility that there were fricative allophones. In fact, Ogura (1998) argues that the fricative sounds would have existed in intervocalic positions. The stop portions of affricates may have weakened between vowels, but remained stable in initial position. Intervocalic weakening may have triggered the further weakening of the intervocalic labial /p/ which was realized as [ɸ], and consequently caused the merger between the intervocalic [ɸ] and the approximant /w/, which has been attested to have taken place around the eleventh century (see section 5 for details). Another account of phonetic realizations has been proposed by Hayata (1977), as mentioned in notes 1 and 10. He argues that /p/, /t/, /k/, and /ts/ were voiced in intervocalic positions, and that these voiced allophones were phonemically distinguished from the prenasalized voiced realizations of /b/, /d/, /g/, and /dz/, respectively (see also Frellesvig 2010: 34–38). These arguments are important for future discussions on this topic.

4.2 Phonetic value and rendaku

Let us now address how sequential voicing, i.e., rendaku, may affect the phonetic value of /ts/. On the basis of the data of rendaku in Old Japanese, Moriyama (1962)

16 Kamei (1973a) discusses various possibilities concerning the hypothetical *s*, and shows the possibility that the initial consonant *ts* of the simplex **tsame* would be sporadically confused with *s* that vanished afterwards.

pointed out that the number of words that undergo rendaku in /ts/ is drastically smaller than the number in any other voiceless segment such as /p/, /t/, and /k/. This discrepancy between /ts/ and the other voiceless obstruents should not be ignored when discussing the phonetic realizations of these phonemes. However, we face a paradoxical problem. If the disparity comes from the difference in the phonetic realization between an affricate and a plain stop, a question arises as to why an affricate is less prone to rendaku than a plain stop. In order to accept the assumed reconstruction of /ts/, it is necessary to explain why only the affricate /ts/ behaves differently from the other voiceless obstruents, despite the fact that voiceless obstruents, including /ts/, have a stop element in common. Furthermore, prenasalization triggered by rendaku should be taken into consideration (note that rendaku is not simple voicing in old Japanese). These problems remain unsolved. The relationship between rendaku voicing and phonetic conditions in old Japanese needs further investigation.

5 Issues on the labial voiceless stop

A quite well-known fact in the history of Japanese is that the phoneme /h/ in modern Japanese comes from the labial sound in old Japanese. The facts involving /p/ or /h/ were taken up quite frequently in textbooks to the extent that it might seem that there would be no room for further consideration. Nevertheless, a number of significant studies have been carried out in the recent decades (Komatsu 1985; Kida 1989; Mori 1991: 97–135; Hayashi 1992; Takeuchi 1995; Ogura 1998; among others). Their concerns include, for example, estimations of phonetic realizations by means of historical documents, investigations into the dialectal variations of relevant sounds, considerations about the dates of the relevant changes, the way in which the changes took place, and factors or conditions under which the relevant changes took place. These subjects are, of course, complicatedly related with each other, and cannot be discussed separately. Since there is a large body of literature about the historical issues on /p/ and /h/, it is impossible to provide a comprehensive survey. We focus here on a few issues closely related to the general phonological studies.

5.1 Changes from old /p/ to modern /h/

Let us briefly sketch the history of this phoneme, addressing important points in chronological order. The consonant system of the earliest stage attested by historical records, including the phoneme /p/, was shown in Table 4 above. In the history, the first change relevant to /p/ is the spirantization in (14). It is still not clear whether

this change took place in every phonological context at once or in intervocalic position before it occurred in word-initial position.

(14) The spirantization in /p/: [p] > [ɸ]

The change in (15) targeted the labial fricative [ɸ] in the intervocalic positions, which merged with the phoneme /w/ around the eleventh century. (16a), (16b), and (16d) illustrate the words involved in this change. The targeted lexical items were native words as in (16a) and (16b) as well as Sino-Japanese morphemes as in (16d). In contrast to these words, the native word in (16c) and the SJ morpheme in (16e) illustrate the forms with the initial /p/, unaffected by the change in (15).

(15) [ɸ] > [w] / V_V

(16)	a.	[kapa] > [kaɸa] > [kawa]	(ModJ. [kawa] 'river')
	b.	[jupu] > [juɸu] > [juu]	(ModJ. [ju:] 'evening')
	c.	[pana] > [ɸana]	(ModJ. [hana] 'flower')
	d.	[kapu] > [kaɸu] > [kau]	(ModJ. [ko:] 'tortoise shell, instep [甲]')
	e.	[pa] > [ɸa]	(ModJ. [ha] 'wave [波]')

As a result of the change in (15), the segmental sequence /Vpu/ merged into /Vu/ (where V indicates a vowel), because phonotactics did not allow /wu/. This is illustrated in (16b) and (16d). All SJ morphemes involved in (15) went along this path, since the intervocalic [ɸ] was followed by no vowels other than /u/. The intervocalic [ɸ] in SJ words corresponds to the coda *p* in classical Chinese. The high vowel /u/ is epenthetic in loanwords, such as [kaɸu] in (16d) that come from *kap* in Classical Chinese (see Kubozono, Ch. 8, this volume for details about epenthetic vowels in loanwords).

The date of the merger in (15) can be collaborated by the confusion observed in writing between *pa, pi, pu, pe, po* and *wa, wi, u, we, wo* (Tsukishima 1969, among others).[17] By contrast, the date of (14) is controversial. Since Hashimoto (1928), it has been generally assumed that /p/ had already been spirantized in or before the Nara period. However, Kida (1989) reexamined the evidence and pointed out the possibility that even if the intervocalic /p/ was spirantized, the initial /p/ may still have been realized as a stop even in the beginning of the Heian period. Hayashi (1992) argued that the bilabial stop [p] remained at least until the ninth century, i.e., the beginning of the Heian period, pointing out that the two changes in (14) and (15) should have occurred in fairly quick succession, because the spirantization

17 [wi], [we], [wo], and [je] are not permitted in the native and SJ words of modern Japanese (Chapters 3 and 8).

in (14) probably naturally triggered the further weakening of stricture in the intervocalic position, and the intervocalic [ɸ] was further loosened and voiced, resulting in the approximant [w] (see also Frellesvig 2010: 34–38).

What is significant in relation to modern Japanese is that the change in (15) brought a phonotactic rearrangement to [ɸ]. The result of the change confined [ɸ] to word-initial positions. Komatsu (1985) interprets this as meaning that this phoneme acquired a demarcative function, which is one of the significant factors behind the change in (15). The phonotactics of this sound established in the eleventh century was transferred over to modern Japanese. In fact, we find /h/ in medial positions in neither the native nor SJ lexicon, except for a few native words such as *ahiru* 'duck' and *ahureru* 'overflow', and a large number of loanwords such as *sohuto* 'soft' borrowed from English.

The third change shown in (17) took place in the period around the second half of the seventeenth century to the eighteenth. As a result of this change, the labial feature vanished from the articulation of this phoneme.

(17) delabialization in /ɸ/(=/h/): [ɸ] > [h]

As for the phonetic realization of /h/, it is quite often pointed out even in some textbooks of Japanese phonetics and phonology that it should be regarded as a devoiced vowel whose quality is the same as that of the vowel that follows it; namely, the sounds should be described as [i̥i], [e̥e], [ḁa], [o̥o], and [ɯ̥ɯ] rather than current IPA representations such as [çi], [he], [ha], [ho], and [ɸɯ]. At any rate, the phoneme /h/ is not straightforwardly specified by any place of articulation. /h/ is an idiosyncratic consonant that is different from other obstruents such as the labial /b/, dental /t/, the velar /k/ in the inventory. The transition from the labial sounds to /h/ as shown in (17) is a remarkable phenomenon in terms of the relationship between phonotactics and sound change in a general sense. The question of how confining /ɸ/ to initial position triggered delabialization has not been substantially discussed in the literature. As to this problem, Kamei et al. ([1976] 2007: 72–87) points out that it took a very long period, i.e., several hundred years, for the initial [ɸ] to change to /h/ after the disappearance of the intervocalic [ɸ] in (15) in the Kyoto dialect, and argued that [ɸ] may have lasted due to surrounding dialects that still preserved the labial realization. He also suggested that the initial bilabial [ɸ] did not change to the labio-dental [f], which is phonetically more stable than [ɸ], under the socio-geographic condition he assumed. Further research is needed, especially concerning how phonotactic properties relate to some sound changes. It is relevant not only to historical studies but also to theoretical considerations.

5.2 Geminate of the labial stop

As mentioned in Kubozono's introduction to this volume and Kawagoe (this volume), the geminate /pp/ (quite often described as /Qp/ in traditional Japanese linguistics)

occurs in modern Japanese. In morphological alternations, this geminate occurs as the counterpart of the singleton /h/ as exemplified in (18).

(18) hai 'cup'
ip-pai (< iti+hai) 'one cup'

The geminate /pp/ occurs quite often in SJ words and occasionally in the native lexicon (*hai* in (18) is a SJ morpheme; see Nasu (this volume) and Ito and Mester (Ch. 7, this volume) for morphophonological aspects of SJ morphemes). Where does the geminate come from, or how did the geminate emerge? In the Japanese Portuguese dictionary *Vocabulario da Lingoa de Iapam* published in Nagasaki in 1603, the geminate is spelled by the roman letters *pp*, similar to common Romanization of modern Japanese. Therefore, the geminate can be dated back at least to the beginning of the seventeenth century, but earlier stages are difficult to verify due to the lack of a distinctive mark in the domestic writing system.

However, we find a few words that may indicate that the geminate /pp/ or the phoneme /p/ with longer duration existed even in an earlier stage, as shown in (19).

(19) a. ModJ. appare 'admirable'
cf. aware 'pathos'
b. ModJ. moppara 'exclusively'

The word *appare* is completely different from the word *aware* in modern Japanese, but the two words are doublets, etymologically derived from a common word. The former originates from the emphasized form of the latter, i.e., *apare*, which was probably pronounced in old Japanese with a longer closure of lips. The form *apare* (or, in the later stage, the spirantized *aɸare*) became *aware* via the merger between intervocalic [ɸ] and [w](=/w/) in (15). While the spirantization and the following merger occurred, the bilabial closure has been preserved in *appare* up to modern Japanese. The merger did not involve *appare*, compared to the other member of the doublet *aware*, and therefore the form with a longer or emphasized /p/ may date back to the stage before the merger that took place around the eleventh century (see section 5.1).

The longer /p/ may have resisted the spirantization due to the solid closure, which may explain why the labial stop in *appare* remains. A similar condition may have been operative in the native word *moppara* in (19) (Kamei et al. [1976] 2007: 82–86; T. Takayama 2002).

5.3 Mimetic p

As often mentioned, the Japanese language has a great number of mimetic expressions, and their role in the lexicon is quite important. The properties of their forms

are also extremely significant from phonological or morphophonological points of view (see Kubozono's introduction to this volume). It is well known that the phonotactics of mimetic forms are different from the phonotactics of plain native words, as discussed in general phonological studies (Ito and Mester 1999, 2008; Nasu 1999; see Nasu, this volume, for details on mimetic phonology). Especially, the initial singleton /p/ in mimetic forms is remarkable (Nasu 1999), and occurs in a large number of mimetic words as shown in (20).

(20) a. piyo piyo 'cheep of chick'
b. pikari(to) 'a flash of light'

Where did the stop /p/ of mimetics come from, if /p/ of Old Japanese changed to the non-labial /h/? The *Vocabulario da Lingoa de Iapam* (1603) lists nine mimetic words (no other words than mimetic) as the entries beginning with /p/. Even in the seventeenth century, both /p/ and /ɸ/ occurred in the consonant system, similar to modern Japanese. The difference between /ɸ/ and /p/ is attested from romanized Japanese as written by Portuguese missionaries (/ɸ/ is represented by the letter *f*, and /p/ by the letter *p*). Directly demonstrating the situation before the seventeenth century is difficult because there was no distinction made in the domestic writing system. However, on the basis of indirect resources, Kamei (1959, 1960) argued that the labial stop probably remained in the mimetic expressions even after the spirantization of /p/. He suggests that the labial voiceless stop was preserved through the history of Japanese since the quality of a sound itself is crucial to mimetic expressions or sound symbolism.[18] Komatsu (1981: 249–283) refers to a kind of morpho-semantic effect quite often observed between voiceless obstruents and voiced (prenasalized) sounds, which is characteristic of Japanese mimetic expressions, as illustrated in (21).

(21) a. pota pota
'with (something like liquid) dripping lightly'

b. bota bota
'with (something like liquid or others) dripping heavily'

Komatsu (1981) argues that since such semantic difference is significant in Japanese sound symbolism, the phonetic parallelism between [p] and [b] should have been preserved in mimetic words, in spite of the spirantization in the plain

18 Kamei (1970b) suggests another historical change involving sound symbolism in Japanese. That is the transition from the affricate /ts/ to a fricative (=/s/ in ModJ), as discussed in section 4.1. He pointed out that the change did not necessarily target mimetic expressions for a reason similar to that seen with the labial stop.

native words, and as a result, that the phoneme /p/ in Old Japanese bifurcated into /ɸ/ and /p/ after the spirantization. Such a split in the history of Japanese provides an important and valuable resource for the discussion of the relationship between sound symbolism and sound changes, both from a general point of view as well as for cross-linguistic investigation of sound symbolism.

6 Conclusion

This chapter first discussed historical issues regarding voiced obstruents in section 2, focusing on the velar nasal [ŋ] and prenasalization in the consonantal system. After introductory remarks on terminology, we dealt with the velar nasal variant [ŋ] of /g/ (section 2.1), which has been given the status of standard in modern Japanese, and looked at the prenasalization that forms the historical background of the velar nasal (section 2.2). Issues concerning the attestation of the prenasalization were also discussed (section 2.3). We returned to the topic of the velar nasal in section 2.4 to discuss some problems in its history. In section 2.5, we dealt with the prehistory of the obstruents in Japanese with focus on the phonotactics of the voiced obstruents and rendaku. Section 3 dealt with the historical background of the affricates, [ts], [tʃ], [dz], and [dʒ], in modern Japanese. Specifically, we looked at two successive sound changes: the affrication of [tu], [ti], [du], and [di], and the mergers between /du/([dzu]) and /zu/([zu]) and between /di/([dʒi]) and /zi/([ʒi]). Furthermore, we looked at the diachronic overlap between the mergers and the loss of prenasalization in the voiced obstruents. Section 4 dealt with issues regarding the phonetic value of the phoneme /s/ at the stages before the affrication of [t] and [d], which were discussed in the previous section. In addition, we also pointed out a dissonance between the reconstructed phonetic value of /s/ and the scarcity of rendaku, i.e., the sequential voicing of /s/ (=/ts/ in Table 4). Reviewing the recent research and the noticeable arguments therein, section 5 presented a history of the phonemes, /h/ and /p/, which resulted from the bifurcation of the labial /p/ in Old Japanese. The historical issues surrounding the geminate /pp/ of modern Japanese as well as those concerning the phoneme /p/ in mimetics were also discussed.

In sum, we have dealt with the main topics of the historical phonology of Japanese in this chapter, especially the consonantal issues that are helpful in understanding the synchronic aspects of modern Japanese. However, there are many important issues that were not discussed in this chapter.

Vowel coalescence is a significant historical event shaping the modern Japanese vowel system and phonotactics. As we saw in section 5.1, the vowel combination /Vu/ occurred in many words, resulting from the merger of intervocalic [ɸ] into /w/ in (15). In addition, there were a large number of SJ morphemes that were not involved in (15) but which originally had /Vu/ arrangements such as *kiu* (九) 'nine', *seu* (少) 'young', *tau* (<*taũ* 唐) 'Tang dynasty', and *you* (<*yoũ* 用) 'use'. Furthermore,

the vowel combination /Vu/ also emerged from other sources (see T. Takayama 1992 for details). Regardless of the source, these combinations have all been replaced by long vowels in modern Japanese. The situation results from the vowel coalescence which is assumed to have ended at the beginning of the seventeenth century. However, modern Japanese has the vowel combination /Vi/. Details on vowel combinations and coalescence are discussed by Kubozono (Ch. 5, this volume) and T. Takayama (1992).

Naturally, the historical background of SJ words is significant for various aspects of the history of Japanese. Phonological problems of SJ words are discussed by Ito and Mester (Ch. 7, this volume) in their relation to modern Japanese.

Finally, there are morphophonological phenomena not addressed here, including so-called "onbin" (Frellesvig 1995). Moreover, the various questions and controversial points concerning the reconstruction of the phonological system of the eighth century were not discussed in this chapter. These problems are dealt with in the History Volume of the same handbook series.

Acknowledgements

I would like to express my special thanks to Donna Erikson for improving the English expressions. I am greatly indebted to Gábor Pintér, Akio Nasu, and Haruo Kubozono for many valuable comments and corrections on an earlier version of this chapter. Any remaining inadequacies and errors are of course my own responsibility.

References

Arisaka, Hideyo. 1936. Jōdai ni okeru sagyō no tōon [/s/ of Old Japanese]. *Kokugo to Kokubungaku* 13(1). Reprinted in Hideyo Arisaka 1957, 145–159.

Arisaka, Hideyo. 1955. Jōdai on'inkō [A study of sounds of Old Japanese]. Tokyo: Sanseido.

Arisaka, Hideyo. 1957. *Kokugo on'inshi no kenkyū: Zōho shinpan* [Studies of phonological history of Japanese: revised version]. Tokyo: Sanseido.

Collado, Didaco. 1632. *Ars grammaticae iaponicae lingae*. Rome. Reproduced in Takanobu Ōtsuka 1957.

Doi, Tadao (trans.). 1955. *Nihon daibunten* [A Japanese translation of Rodriguez 1604–1608]. Tokyo: Sanseido.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Fukushima, Kunimichi (ed.). 1989. *Rodorigesu nihon shōbunten* (A reproduction of Rodriguez. 1620, in the collection of the school of Oriental and African studies of London University). Tokyo: Kasama Shoin.

Frellesvig, Bjarke. 1995. *A case study in diachronic phonology: The Japanese Onbin sound changes*. Aarhus: Aarhus University Press.

Frellesvig, Bjarke. 2010. *A history of the Japanese language*. New York: Cambridge University Press.

Hashimoto, Shinkichi. 1928. Hagyō shiin no hensen ni tsuite [A history of the voiceless labial consonant]. In Sanki Ichikawa (ed.), *Okakura sensei kinen ronbunshū* [Festschrift of professor Okakura]. Reprinted in Shinkichi Hashimoto 1950, 30–45.

Hashimoto, Shinkichi. 1932. Kokugo ni okeru biboin [Nasal vowels in Japanese]. *Hōgen* 2(1). Reprinted in Shinkichi Hashimoto 1950, 1–9.

Hashimoto, Shinkichi. 1950. *Kokugo on'in no kenkyū* [Studies on Japanese phonology]. Tokyo: Iwanami.

Hattori, Shiro. 1955. On'inron (1) [Phonology (1)]. *Kokugogaku* 22. Reprinted in Shiro Hattori 1960, 279–301.

Hattori, Shiro. 1956. On'inron (2) [Phonology (2)]. *Kokugogaku* 26. Reprinted in Shiro Hattori 1960, 302–322.

Hattori, Shiro. 1957. On'inron (3) [Phonology (3)]. *Kokugogaku* 29. Reprinted in Shiro Hattori 1960, 323–352.

Hattori, Shiro. 1960. *Gengogaku no hōhō* [Methodologies in linguistics]. Tokyo: Iwanami.

Hayashi, Chikafumi. 1992. Hagyō tenko on wa naze Heian jidai ni okotta ka: Nihongo on'inshi no shiten to kijutsu [Why did the intervocalic ɸ merge into w in the Heian period?]. *Kokugo to Kokubungaku* 69(11). 110–119.

Hayashi, Chikafumi. 2002. Kyū-seiki nihongo no shiin onka: Nihongo on'inshi ni okeru bunkengaku-teki kōsatsu no imi to hōhō [Phonetic values of the consonants of Japanese in the 9th century]. *Kokugo to Kokubungaku* 78(4). 1–14.

Hayata, Teruhiro. 1977. Seisei akusentoron [Generative analysis of pitch accent]. In Susumu Ōno and Takesi Sibata (eds.), *Iwanami kōza nihongo 5: On'in* [Iwanami Japanese series 5: Phonology], 323–360. Tokyo: Iwanami.

Hibiya, Junko. 1988. Bariēshon riron [Variation theory]. *Gengo Kenkyu* 93. 155–171.

Hibiya, Junko. 1995. The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change* 7(2). 139–152. Reprinted in Natsuko Tsujimura 2005.

Hibiya, Junko. 2002. Gengo hen'i no chiriteki sai [Regional differentiation of Linguistic Variation]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 6(3). 60–68.

Hino, Hiroshi. 1993. *Nihon shōbunten* [A Japanese translation and reproduction of Rodriguez. 1620, in the collection of the Ajuda library, Lisbon]. Tokyo: Shinjinbutsu Ōraisha.

Hizume, Shūji. 2003. Seidaku bunka to sokuon, hatsuon [The establishment of *sei-daku* opposition and moraic obstruents/nasals]. *Kokugogaku* 54(2). 95–108.

Inoue, Fumio. 1971. Gagyōshiin no rekishi [The distribution and history of the g-sound in Japanese]. *Kokugogaku* 86. 20–41.

Inoue, Fumio (ed.). 1983. *Shinhōgen to kotoba no midare ni kansuru shakai gengogakuteki kenkyū: Tokyo, Shutoken, Yamagata, Hokkaido* [A sociolinguistic study of "new dialect" and "language deterioration": Tokyo, Metropolitan area, Yamagata, Hokkaido]. (KAKENHI report).

Inoue, Fumio. 1998. *Nihongo uocchingu* [Watching Japanese]. Tokyo: Iwanami.

Ito, Junko and R. Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17(1). 49–73.

Ito, Junko and R. Armin Mester. 1997. Correspondence and compositionality: The ga-gyō variation in Japanese phonology. In Iggy M. Roca (ed.), *Derivations and constraints in phonology*, 419–462. New York: Oxford University Press.

Ito, Junko and R. Armin Mester. 1999. The structure of the phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Malden, MA, and Oxford: Blackwell.

Ito, Junko and R. Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure.* Cambridge, MA: MIT Press.

Ito, Junko and R. Armin Mester. 2008. Lexical classes in phonology. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 84–106. Oxford: Oxford University Press.

Ito, Junko and R. Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Itoi, Kanichi. 1962. Ōita-ken no yotsugana ben ryakuhō: Kusu-machi Kimuta hōgen ni tsuite [A report on the contrasts of *zi* and *di*, *zu* and *du* in the Kusu machi Kimuta dialect of Ōita prefecture]. *Kokubungakukō* 27. 48–52.

Kamei, Takashi. 1950. Kenshukuryōkoshū o chūshin ni mita yotsugana [On the contrasts of *zi* and *di*, *zu* and *du* in *Kensyukuryōkosyū*]. *Kokugogaku* 4. Reprinted in Takashi Kamei 1984, 291–312.

Kamei, Takashi. 1956. Gagyō no kana [On kana letters for the syllables with the g-sound]. *Kokugo to Kokubungaku* 33(9). Reprinted in Takashi Kamei 1984, 1–25.

Kamei, Takashi. 1959. Shunnōten. *Kokugogaku* 39. Reprinted in Takashi Kamei 1984, 141–153.

Kamei, Takashi. 1960. Kodaijin no warai goe [Laughter of ancient people]. *Nihon Koten Bungaku Taikei Geppō* 37. Reprinted in Takashi Kamei 1984, 371–375.

Kamei, Takashi. 1970a. Kana wa naze dakuon sen'yō no jitai o motanakatta ka o megutte kataru [On the reasons why the kana system does not have letters for daku-on]. *Jinbun Kagaku Kenkyū* 12. Reprinted in Takashi Kamei 1986, 63–150.

Kamei, Takashi. 1970b. Suzume shiu shiu [Sparrows chirp 'siu siu']. *Seikei Kokubun* 3. Reprinted in Takashi Kamei 1984, 447–464.

Kamei, Takashi. 1973a. Bunken izen no jidai no nihongo [Prehistoric Japanese]. Reprinted in Takashi Kamei 1973b, 265–399.

Kamei, Takashi. 1973b. *Nihongo keitōron no michi (Kamei Takashi Ronbunshū 2)* [Perspectives on the cognation between Japanese and other languages (Selected works of Kamei Takashi vol. 2)], 265–399. Tokyo: Yoshikawa Kōbunkan.

Kamei, Takashi. 1984. *Nihongo no sugata to kokoro* 1 *(Kamei Takashi Ronbunshū 3)* [Meaning and form in Japanese 1 (Selected works of Kamei Takashi vol. 3)]. Tokyo: Yoshikawa Kōbunkan.

Kamei, Takashi. 1986. *Gengo bunka kusagusa (Kamei Takashi Ronbunshū 5)* [Considerations on Language and Culture (Selected works of Kamei Takashi vol. 5)]. Tokyo: Yoshikawa Kōbunkan.

Kamei, Takashi, Tokihiko Ōtō and Toshio Yamada (eds.). 2007 [1976]. *Nihongo no rekishi 5: Kindaigo no nagare* [History of the Japanese Language vol. 5: Trends in Modern Japanese], 2nd edition. Tokyo: Heibonsha.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kida, Akiyoshi. 1989. P'on zokukō [Further consideration on the p sound]. In Okumura Mitsuo kyōju taikan kinen ronbunshū kankōkai (eds.), Okumura Mitsuo kyōju taikan kinen ronsō [Festschrift in celebration of the retirement of Professor Mitsuo Okumura], 415–429. Tokyo: Ōfūsha.

Kim, Hyunsoon. 2001. A phonetically based account of phonological stop assibilation. *Phonology* 18 (1). 81–108.

Kindaichi, Haruhiko. 1942. Gagyō bionron [On the Japanese velar nasal]. Reprinted in Haruhiko Kindaichi 1967, 168–197.

Kindaichi, Haruhiko. 1967. *Nihongo on'in no kenkyū* [Studies in Japanese Phonology]. Tokyo: Tōkyōdō Shuppan. Reprinted in Haruhiko Kindaichi 2005.

Kindaichi, Haruhiko. 2005. *Kindaichi Haruhiko chosakushū 6* [Selected works of Haruhiko Kindaichi vol. 6]. Tokyo: Tamagawa University Press.

Komatsu, Hideo. 1981. *Nihongo no sekai 7: Nihongo no on'in* [The world of Japanese 7: Japanese phonology]. Tokyo: Chūōkōronsha.

Komatsu, Hideo. 1985. Bogo no rekishi o toraeru shiten [A viewpoint to deal with the history of native language]. In Shirō Hayashi (ed.), *Ōyō gengogaku kōza 1: Nihongo no kyōiku* [Series in applied linguistics 1: Japanese teaching], 132–147. Tokyo: Meiji Shoin.

Kubozono, Haruo. 2012. Varieties of pitch accent systems in Japanese. *Lingua* 122(13). 1395–1414.
Kubozono, Haruo. this volume. Introduction to Japanese phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kubozono, Haruo. this volume. Chapter 5: Diphthongs and vowel coalescence. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Kuno, Mariko, Makoto Kuno, Makio Ōno and Takao Sugimura. 1995. Yotsugana tairitsu no shōshutsu katei: Kōchi-ken Nakamura-shi, Aki-shi no baai [The mergers of *zi* and *di*, *zu* and *du* in the Nakamura and Aki dialects of Kōchi Prefecture]. *Kokugogaku* 180. 84–96.
Kuno, Makoto. 2006. Kōchi hōgen no zenbion [Pre-nasalization of voiced plosives in Kochi dialect]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 10(1). 60–69.
Lin, Yen-Hwei. 2011. Affricates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 1*, 367–390. Malden, MA and Oxford: Wiley-Blackwell.
Mori, Hiromichi. 1991. *Kodai no on'in to Nihon shoki no seiritsu* [Sounds of Old Japanese and production of *Nihon shoki*]. Tokyo: Taishukan.
Morita, Takeshi. 1977. Nippo jisho ni mieru go'on renketsu jō no ichi keikō [A combinatorial tendency of sounds in *Vocabulario da Lingoa de Iapam*]. *Kokugogaku* 108. 20–32.
Moriyama, Takashi. 1962. Rendaku: Jōdai ni okeru [Rendaku in Old Japanese]. *Gobun Kenkyū* 14. Reprinted in Takashi Moriyama 1976, 279–297.
Moriyama, Takashi. 1976. Jōdai kokugo on'in no kenkyū [Studies on sounds in Old Japanese]. Tokyo: Ōfūsha.
Nasu, Akio. 1999. Onomatope ni okeru yūseika to [p] no yūhyōsei [Voicing in onomatopoeia and the markedness of [p]]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 3(3). 52–66.
Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Ogura, Hajime. 1998. Sagyō shiin no rekishi [Ancient Japanese sibilants]. *Kokugogaku* 195. Reprinted in Hajime Ogura 2011, 137–160.
Ogura, Hajime. 2011. *Nihongo on'inshi ronkō* [Research on the history of Japanese phonology]. Osaka: Izumi Shoin.
Ōtsuka, Takanobu. 1957. *Koryādo nihon bunten* [A Japanese translation and a reproduction of *Ars grammaticae iaponicae linguae*]. Tokyo: Kazama Shobō.
Pintér, Gábor. this volume. Chapter 3: The emergence of new consonant contrasts. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
Riehl, Anastasia K. and Abrigail C. Cohn. 2011. Partially nasal segments. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology 1*, 550–576. Malden, MA and Oxford: Wiley-Blackwell.
Rodriguez, João. 1604–1608. *Arte da lingoa de Iapam*. Nagasaki. Japanese annotation by Tadao Doi, 1976. *Nihon bunten*. Tokyo: Benseisha.
Rodriguez, João. 1620. *Arte breve da lingoa de Japoa*. Macao. Reproduced by Kunimichi Fukushima 1989 and reproduced with Japanese translation by Hiroshi Hino 1993.
Sakanashi, Ryūzō. 1975. Sanba no shiroki nigori ni tsuite [On the diacritical circle of Sanba]. *Okayama Daigaku Hōbungakubu Gakujutsu Kiyō* 36. 142–127. Reprinted in Ryūzō Sakanashi 2004, 253–280.
Sakanashi, Ryūzō. 2004. *Kinsei no goi hyōki* [Aspects of spellings in the Edo period]. Tokyo: Musashino Shoin.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie K. Huffman and Rena A. Krakow (eds.), *Nasals, nasalization, and the velum* (*Phonetics and phonology* 5), 401–470. San Diego: Academic Press.

Sugimura, Takao. 2001. Kyūshū hōgen no yotsugana [On the contrasts of *zi* and *di*, *zu* and *du* in the phonemic systems of the Kyushu dialects]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 5(3). 10–18.

Takayama, Michiaki. 1992. Seidaku shōkō [A consideration on sei-daku]. In Ikudō Tajima and Kazuya Niwa (eds.), *Nihongo ronkyū 2: Koten nihongo to jisho* [Studies of Japanese vol. 2: Classical Japanese and dictionaries], 17–56. Osaka: Izumi Shoin.

Takayama, Michiaki. 2006. Yotsugana to zenbion [Prenasalization and merger]. In Chikushi kokugogaku danwakai (ed.), *Chikushi gogaku ronsō II: History of Japanese and dialects*, 158–174. Tokyo: Kazama Shobō.

Takayama, Michiaki. 2012. *Nihongo on'inshi no kenkyū* [A study of phonological history of Japanese]. Tokyo: Hituzi Syobo.

Takayama, Tomoaki. 1992. Nihongo ni okeru rensetsu boin no chōboinka: Sono rekishiteki imi to hassei no onseiteki jōken [The changes of Vu hiatuses into long vowels in the history of Japanese language]. *Gengo Kenkyu* 101. 14–34.

Takayama, Tomoaki. 1993. Hasatsuon to masatsuon no gōryū to dakushiin no henka: Iwayuru yotsugana gōryū no rekishiteki ichizuke [Transition and merger in the voiced obstruents of late middle Japanese]. *Kokugo Kokubun* 62(4). 18–30.

Takayama, Tomoaki. 1998. 17 seiki matsu no zenbion no jittai ni tsuite: Ikeisai kōgo kikigaki and Waji shōran shō no saikenshō [On the prenasalization in the end of the seventeenth century: Re-examinations of *Ikeisai kōgo kikigaki* and *Waji shōran shō*]. *Kagawa Daigaku Kokubun Kenkyū* 23. 1–8.

Takayama, Tomoaki. 2002. Nihon kango no shiteki on'inronteki kadai [Some issues in historical phonology of Sino-Japanese]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 6(1). 44–52.

Takayama, Tomoaki. 2003. Kenshukuryōkoshū kara ukagaeru zenbion'yōso no ichi kyokumen [An aspect of the de-prenasalization seen in *Kenshukuryōkoshū*]. *Kagawa Daigaku Kokubun Kenkyū* 28. 94–101.

Takayama, Tomoaki. 2006. Hasatsuonka to bointaikei [Affrication and vowel system], In Jōo Hakutarō hakushi kanreki kinen ronbunshū henshū iinkai (ed.), *Jikken onseigaku to ippan gengogaku: Jōo Hakutarō hakushi kanreki kinen ronbunshū* [Experimental phonetics and general linguistics: Festschrift in celebration of the 60th anniversary of Dr. Jōo Hakutarō], 218–228. Tokyo: Tōkyōdō Shuppan.

Takayama, Tomoaki. 2009. Tagyō dagyō hasatsuonka no on'inronteki tokushitsu [Contrastive aspects of the affrications in late middle Japanese]. *Kanazawa Daigaku Kokugo Kokubun* 34. 203–215.

Takayama, Tomoaki. 2010. Yotsugana konran to zenbishiin suitaika to no kanshō: Kobetsu gengoshi no shiten no jūyōsei [Accidental synchronization between de-prenasalization and merger]. In Tsukimoto, Masayuki, Toshihiro Fujii and Shūji Hizume (eds.), *Kotengo kenkyū no shōten*, 851–871. Tokyo: Musashino Shoin.

Takeuchi, Nobuo. 1995. Yōshū oyobi chōan ni okeru Ennin no shittan gakushū: Tokuni Hōgetsu 'shiddan jibo' o megutte [Ennin's learning of siddham in Yang-zhou and Chang-an and his note on the pronunciation of siddha-mātrikā letters]. *Tokyo Daigaku Kyōyōgakubu Kiyō: Hikaku Bunka Kenkyū* 33. 1–146.

Tsujimura, Natsuko (ed.). 2005. *Japanese linguistics* 3. *Pragmatics, sociolinguistics, and language contact*, 1–14. New York: Routledge.

Tsukishima, Hiroshi. 1969. *Heian jidaigo shinron* [A novel approach to Japanese of the Heian period]. Tokyo: Tokyo University Press.

Uwano, Zendo. 2012. Three types of accent kernels in Japanese. *Lingua* 122(13). 1415–1440.
Vance, Timothy J. 1987. *An Introduction to Japanese phonology.* Albany, NY: State University of New York Press.
Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.
Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.
van de Weijer, Jeroen M., Kensuke Nanjo and Tetsuo Nishihara (eds.). 2005. *Voicing in Japanese*. Berlin and New York: Mouton de Gruyter.
Yamaguchi, Yoshinori. 1988. Kodaigo no fukugōgo ni kansuru ichi kōsatsu [On compounds in Old Japanese]. *Nihongogaku* 7(5). 4–12.
Yamane-Tanaka, Noriko. 2005. The implicational distribution of prenasalized stops in Japanese. In Jeroen M. van de Weijer, Kensuke Nanjo and Tetsuo Nishihara (eds.), *Voicing in Japanese*, 123–156. Berlin and New York: Mouton de Gruyter.

Kikuo Maekawa

16 Corpus-based phonetics

1 Introduction: Notes on some key notions about speech corpora

The aim of the present chapter consists in giving an overview of the works done in the field of Japanese phonetics and phonology that involve corpus analysis as their essential ingredient. As such, this paper consists of two parts: one that provides information about Japanese corpora, and the other that provides information about the achievements of corpus-based studies. The first and last halves of this chapter will be devoted to these respective goals.

The description of this chapter is as follows. The rest of section 1 is devoted to the explanation of some key notions in this field. In section 2, some publicly available Japanese speech corpora are introduced. Section 3 is the main body of the chapter, which begins with an introduction to the early corpus-based studies done by Japanese researchers in the 1950s and 1960s. The rest of the section is devoted to summaries of the main achievements in this field in the present time. They are classified in six subfields, viz., the analysis of segment duration, the analysis of speaking style and phonetic variations, the analysis of filled pauses, the analysis of dialogue and discourse, the analysis of infant speech and infant-directed speech, and the analysis of paralinguistic and extra-linguistic information. Section 4 is a concluding section that provides an overall recapitulation of the trends of corpus-based studies in Japanese phonetics and phonology as observed in the studies presented in section 3. Some problems with these studies, both apparent and potential, are also pointed out.

Before introducing Japanese speech corpora, a digression is required in order to avoid unnecessary misunderstandings stemming from inconsistent usage of some key notions in the field.

Corpus, to begin with, is one such notion. Recently, many linguists and phoneticians use this word as a mere synonym of *data*. The usage is awkward, because the word ‘corpus’ was originally introduced to make reference to a special kind of linguistic data to the exclusion of all others. First, the data should have *authenticity* of some kind. The language samples in a corpus must record real language behavior. The notion of authentic data is the opposite of that of artificial data. Data gathered in experimental settings, as well as data generated based on introspection can hardly be regarded as corpus data in the original sense of the word. It is to be noted here that if we make a strict interpretation of authenticity as a prerequisite of corpus, many of the ‘corpora’ created in the field of speech engineering are not corpus data anymore, because they are recorded in artificial settings.

Second, the data should have *representativeness* of some sort. This means that a corpus should be an exact epitome of the target language variety. Here, it is important to note that, in the case of spoken corpora, it is far more difficult to guarantee corpus representativeness than in the case of written corpora. Unlike written language which is inherently time-resisting, speech materials disappear as soon as they are generated. Although devices to make recordings of spoken material are wide-spread in contemporary society, it is nonetheless the case that the greater part of spoken language materials is lost as soon as the spoken material is produced. The consequence is that it is virtually impossible, in an overwhelming majority of cases, to think about the population of authentic speech behavior from which the corpus samples are to be selected. Designers of speech corpora pay much attention to factors like properties of speakers (age, gender, birth place, education level, etc.), mode of speech (monologue, dialogue, multi-party conversation, etc.), and speech registers (academic presentation, public speech, interview, casual talk, reading, etc.) so that the corpora possess some sort of representativeness. Still, it is true that speech corpora have less representativeness compared to carefully designed written corpora.

Third, the coverage of speech registers is often referred to as the issue of *corpus balance*. Here again, spoken language is much less clear than written language in terms of our knowledge about the total range of possible registers.

At this point, it is to be noted that there is another kind of balance that is discussed with respect to speech corpora, i.e., *phoneme-balance*. A phoneme-balanced corpus is the one in which all phonemes or all short combinations of phonemes (like, CV or CVC) of the target language are included. Phoneme-balance is of particular importance when a corpus is applied to make a language-model for automatic speech recognition and other speech information processing purposes.

The simplest form of a speech corpus would be the combination of recorded speech and its transcription. There are, however, spoken language corpora that consist exclusively of transcription texts. The renowned *London-Lund Corpus* is an example. The availability of such corpora as resources for phonetic research is considerably limited compared to corpora that are distributed with audio recordings.

The notion of annotation differs considerably from one corpus to another. Maekawa (2013a) notes that transcription text should be regarded as an annotation, implying that the essential main body of a speech corpus is the audio signal per se. According to his interpretation, the above-mentioned speech corpora consisting of only transcription text should be regarded as corpora consisting only of annotation data.

Annotation of speech corpora can be divided into two categories: linguistic and phonetic. Typical linguistic annotations involve morphological and syntactic information as applied to transcription texts. Phonetic annotations are either segmental or prosodic.

Systems of prosodic annotation differ considerably depending on the purposes of the annotation. The simplest prosodic annotation is the indication of highly ear-catching phrase-final local pitch movements (boundary pitch movements, or BPM, see Venditti, Maekawa, and Beckman 2008), which are symbolized by various arrow-like symbols and inserted into the transcription texts. The transcription of the London-Lund Corpus is a typical example.

On the other hand, there are highly systematic prosodic annotation schemes like the J_ToBI and X-JToBI systems (Venditti 1997; Maekawa et al. 2002; Igarashi, Kikuchi, and Maekawa 2006) that are applied to the audio signal rather than the transcriptions. As will be shown in subsections of section 3 below, prosodic annotation like X-JToBI is quite useful for phonetic/linguistic analyses of spontaneous speech, but its application is very time-consuming, hence relatively exceptional.

Lastly, from a historical point of view, not all speech corpora are computerized. Studies based upon the corpora of the pre-computer age will be referred to in section 3.1 below.

2 Japanese speech corpora

This section gives a brief overview of Japanese speech corpora. Table 1 shows some of the major Japanese speech corpora, which are classified according to (a) age of speakers (adult versus infant), (b) availability of audio signal, (c) spontaneity, and, (d) the mode of speech (monologue versus dialogue).

Table 1: Some of publicly available Japanese speech corpora

AGE OF SPEAKERS	AUDIO	SPONTANEITY	MONOLOGUE	DIALOGUE
Adult	Yes	Spontaneous	*CSJ (APS and SPS)*	*Chiba Map Task Corpus* *CALLHOME Japanese* *Utsunomiya University Paralanguage Corpus* *CSJ* (Dialogue) *Nihongo Gakushūsha Kaiwa DB*
		Reading	*JNAS (Newspaper), ATR(Phoneme-balance), CSJ (Reproduction)*	*ASJ-JIPDEC* (navigation task)
	No	Spontaneous	*Minutes of National Diet*	*KY Corpus* *Meidai Kaiwa Corpus* *C-JAS*
Infant	Yes Yes/No	Spontaneous		*NTT Infant Speech DB* *CHILDES corpora*

Note that Table 1 is not an exhaustive list of existing speech corpora; there are many more speech corpora whose target language is Japanese. Readers can obtain information from the web sites of institutions like NII-SRC (http://research.nii.ac.jp/src/en/index.html), GSK (http://www.gsk.or.jp/index_e.html), ALAGIN (http://www.alagin.jp/index-e.html), ATR (http://www.atr-p.com/qhm/), and NINJAL (http://www.ninjal.ac.jp/corpus_center/). In addition, Itahashi and Tseng (2010) provides a wider view of the details of speech corpora in Japanese and other Asian languages.

2.1 Monologue corpora

As shown in the table, monologue corpora can be divided into three types: corpus of spontaneous speech with audio signal, corpus of read speech with audio signal, and, corpus of spontaneous speech without audio signal.

As for the first type, CSJ, or *Corpus of Spontaneous Japanese* (Maekawa et al. 2000; Maekawa 2003, 2004b) is the most typical. It was developed primarily as data for the construction of acoustic- and language-models required for the development of an ASR (automatic speech recognition) system that could handle speech materials that were more or less spontaneous, but it was also designed for phonetic and/or linguistic studies of spontaneous speech. Special annotations were given to a special subset of the CSJ. See below.

The CSJ includes 652 hours of speech materials corresponding to about 7.5 million words uttered by 1,472 different speakers. 95% of the corpus is devoted to spontaneous monologue (see below for the remaining 5%) consisting of two speech registers: APS (academic presentation speech; a live recording of presentations done in 11 academic societies of engineering, social sciences, and humanities) and SPS (simulated public speaking; public speaking done by recruited laymen subjects on everyday topics like 'the most joyful/saddest memory of my life', 'the town where I live' and so forth).

All materials of the CSJ are finely transcribed with time-alignment information. The transcription texts are fully analyzed in terms of POS (part-of-speech) information and CBL (clause boundary labeling) information. Moreover, 44 hours (half a million words) of CSJ are annotated with respect to segmental and prosodic properties using the X-JToBI annotation scheme mentioned above. This subset of the corpus is called the CSJ-Core.

As for the second type (read monologue), these are corpora developed for the study of ASR (automatic speech recognition) and speech synthesis. JNAS is a corpus developed by the Acoustical Society of Japan (Itou et al. 1999). It consists of readings of 16,176 newspaper articles (about 60 hours) and a list of 503 phoneme-balanced sentences by 306 speakers (153 males and females).

ATR laboratories delivers various speech corpora (Sagisaka and Uratani 1992), but it is probably the *Set B* database that is used most widely. The database consists

of the reading of a phoneme-balanced selection of 10,000 sentences that are extracted randomly from written registers covering newspapers, magazines, novels, letters, textbooks and so forth, read by ten speakers. The audio signal is annotated with respect to phonemic segments and prosodic events.

A part of CSJ (sixteen talks of about four hours long) is devoted to a special type of read speech known as *reproduction* speech. This is the reading aloud of the transcribed spontaneous speech (APS and SPS) by the same speakers who produced the original spontaneous speech.

As for the third type (spontaneous monologue without audio signal), only one corpus belongs to this type, i.e., the *Minutes of the National Diet*, or *Kokkai Kaigiroku*. This is an archive of Japan's National Diet covering all meetings held after 1945 (http://kokkai.ndl.go.jp/). This corpus is so-called *monitor corpus* in that it is continuously expanding and used as a language corpus mainly by linguists who are interested in the ongoing linguistic changes in present-day Japanese. Although the lack of audio signal imposes serious limitations on the applicability of the corpus, it is still possible to use the corpus as a resource for phonological studies.

2.2 Dialogue corpora

In contrast to monologue, there are several corpora in the category of spontaneous dialogue. Corpora in this category have been developed for various research purposes. The *Chiba University Japanese Map Task Dialogue Corpus* (Horiuchi et al. 1999), also known as *The Japanese Map Task Dialogue Corpus*, is a Japanese counterpart of the *HCRC Map Task Corpus* (Anderson et al. 1991). Just like the original, this is basically a corpus of task-oriented dialogue between an information giver and a follower. It consists of 128 dialogues of about 23 hours done by 64 speakers. The corpus consists of audio signals and transcriptions. The transcription is written exclusively in *hiragana*. As was the case with the original English version, the corpus is constructed for dialogue analysis research and other psycholinguistic research purposes.

The *CALLHOME Japanese Corpus* consists of 120 spontaneous telephone conversations by native speakers. This corpus was developed originally for the ASR of conversational speech, and is a part of the larger CALLHOME corpus series that covers a variety of languages including English, Arabic, Chinese, and Spanish, among others. The corpus consists of audio signals (maximum size is 30 minutes long) and their transcriptions, but the transcriptions do not cover the whole body of the corpus; transcriptions 10 minutes long are provided for 100 conversations, and, transcriptions 5 minutes long are provided for the rest. Substantial parts of the corpus are left untranscribed. See Den and Fry (2000) for their attempt to annotate the corpus with morphological, prosodic, and semantic information.

The *Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies* (UUDB) is similar to the Map Task Corpus in that it is also a corpus of task-oriented dialogue; the task of the UUDB is the reconstruction of a four-panel cartoon from the data of randomized panels. The audio signal is transcribed and annotated with respect to so-called paralinguistic information labels representing the six basic dimensions of the speakers emotional states: 'pleasant–unpleasant', 'aroused–sleepy', 'dominant–submissive', 'credible–doubtful', 'interested–indifferent', and, 'positive–negative' (Mori et al. 2011). The corpus consists of seven dialogues done by 12 female and two male speakers and includes 4,737 utterances.

A part of the CSJ is devoted to dialogue speech. The dialogue data consists of interviews on the contents of APS and/or SPS talks, task-oriented dialogues, and free dialogues by the speakers of spontaneous monologues. Using the CSJ, it is possible to make three-way comparison among the monologue (APS and SPS), dialogue, and reproduction speeches, but the number of speakers who provided all these speeches is limited (eight males and eight females).

Lastly, there is a corpus of Japanese learners' interview speech. The *Nihongo Gakushūsha Kaiwa Database* (https://dbms.ninjal.ac.jp/nknet/ndata/opi/). This is a collection of the transcriptions of 339 OPI interviews. OPI is the abbreviation for Oral Proficiency Interview, which is a face-to-face interview conducted to estimate the foreign language proficiency level of the interviewee. Audio signal is also available for a subset (215 interviews) of the corpus. The native languages of the interviewees include Korean (207), Chinese (66), English (30), Indonesian (14), and so forth.

The second type of dialogue corpora in Table 1 is a corpus of read dialogue speech, i.e., the *ASJ Continuous Speech Corpus for Research* (ASJ-JIPDEC), where 'ASJ' stands for the Acoustical Society of Japan. Part of this corpus is devoted to the reading of transcribed dialogue by multiple speakers. Dialogue speech of various navigation tasks was recorded and transcribed. The transcribed texts were edited for reading (for example, all filled pauses are removed from the transcription) and read aloud by 36 (18 male and 18 female) speakers. This subcorpus was developed mainly for the study of ASR of dialogue speech. See Kobayashi et al. (1992) for details.

The third type of dialogue corpus consists of only transcription texts. The first two corpora belonging to this type in Table 1, the *KY Corpus* and the *Meidai Kaiwa Corpus*, were developed to obtain research material for teaching Japanese as a foreign language, or TJFL.

The *KY corpus* (http://www.opi.jp/shiryo/ky_corp.html) consists of transcriptions of OPI interview material of 90 Japanese language learners speaking Chinese, Korean, and English as their native languages (30 learners for each language). The proficiency levels of the speakers of the corpus distribute across 'novice', 'intermediate', 'advanced', and 'superior'.

The *Meidai Kaiwa Corpus* consists of transcriptions of about 100 hours of spontaneous conversations of Japanese native speakers (https://dbms.ninjal.ac.jp/nknet/ndata/nuc/). Most of the samples are dialogues, but there are also some multi-party conversations. The transcription includes simple annotations of rising intonation, laughter, pauses, back channels, and inaudible segments.

The third corpus is the *Corpus of Japanese as a Second Language* (C-JAS; https://ninjal-sakoda.sakura.ne.jp/c-jas/web/), which is a longitudinal record of six learners (three Chinese native speakers and three Korean native speakers) covering three years with a sampling interval of 3-4 months. The size of the corpus is about 800 thousand words.

The fourth, and last type of dialogue corpus is concerned with the spontaneous speech of infants and/or their parents. The *NTT Infant Speech Database* consists of recordings of utterances produced by five infants of three families and their parents. Recordings of one hour long were conducted once every month from the infants' births till they became five years old. The corpus provides phonetic information (F0 information and voiced/unvoiced flags) in addition to audio signals and transcriptions (Amano et al. 2009).

Lastly, CHILDES, or child language data exchange system, is a cover term for various speech data concerning mostly native language acquisition and, to a lesser extent, L2 learning registered under the rubric of CHILDES. As far as Japanese is concerned, there are a dozen corpora differing in content, size, and annotation. Audio signal is available for some of them. See Miyata (2004) for more details.

3 Achievements of corpus-based studies

3.1 Works in the pre-computerized era

Nowadays, being digitized is one of the basic requirements for a corpus; corpora listed in Table 1 are all digitized corpora. There were times, however, when linguistic studies were conducted using non-digitized corpora. The contributions of such studies should not be underestimated. Three important books about corpus-based studies of spontaneous Japanese were written by researchers of the National Language Research Institute (NLRI) using pre-computerized corpora that they built. NLRI is the predecessor institution of the current National Institute for Japanese Language and Linguistics (NINJAL), or *Kokuritsu Kokugo Kenkyūjo*.

The first book of the series was entitled *Danwago no jittai* and was published in 1955 (NLRI 1955). The aim of the book was to conduct exploratory research concerning basic issues of spoken language, including 'intonation; length of words, *bunsetsu*, and sentences; structure of sentences; and the type, frequency, and usage of words' (p. 1, translation mine) by use of transcriptions of everyday speech as captured by magnetic tape-recorder.

Materials of daily conversations were recorded in various places in Tokyo, and annotated with respect to the speakers' properties and the social settings of the conversations. The former included sex, age, and education-level, and the latter included dialect area in Tokyo (*Yamanote* versus *Shitamachi*), places where the conversation took place (home, school, work place, etc.), number of parties, and, personal relationship among the parties (known versus unknown).

Recordings were also made of radio news, news commentary, *rakugo* (traditional comic storytelling), lectures, and theatrical performances for the sake of comparison with the daily conversations. 77 reels of magnetic tape were recorded in the years 1953–54, the total amount of which is estimated to be about 30 hours.

Below, the content of chapter 2 of the book, which was devoted for the description of intonation in spontaneous Japanese, will be summarized. In this chapter, the phrase- and sentence-final pitch shapes, like 31 (rising), 33 (level), or 23 (falling), were described using a four-level pitch description system (where 1 is the highest and 4 is the lowest).

The main findings include the following: (a) The most frequent pitch shapes are 33 and 32, while 22, 23, 31, and 21 are the second most frequent. (b) No substantial difference could be found between the sentence-final and phrase-final locations with respect to the inventories and the frequency distributions of pitch shapes. (c) Influence of sex, age, or education level upon the choice of intonation was not clearly observed. However, (d) clear difference was observed between the intonations of radio-news and daily conversation (e.g., pitch shape 32 appeared frequently in daily conversation, while it appeared only once in the radio-news data).

No matter how primitive it may appear from today's perspective, the book and the chapter was the first light of dawn in the history of corpus-based studies of spoken Japanese and the phonetics of spontaneous Japanese. Perhaps, it is worth noting here that this is not only the first systematic study of spontaneous speech of the Japanese language, but also one of the earliest corpus-based studies of spoken language in the world.[1]

In subsequent years, corpus-based study of spoken language was continued by the researchers of the NLRI along the lines developed by *Danwago no jittai*. The results were reported in two volumes of *Hanashi kotoba no bunkei* (NLRI 1960, 1963). The first and second volumes were devoted respectively to analyses of dialogue and monologue materials. As far as Japanese is concerned, there is wide consensus among researchers of Japanese linguistics and phonetics that these two volumes established a firm basis for the study of spoken language.

The analytical topics covered by the two volumes include (a) definition of 'sentence' in spoken language, (b) grammatical properties of spoken language (covering the minor topics like mood, voice, tense, and the various means to express

---

1 The NLRI study was much in advance of the renowned *Survey of English Usage* by Randolph Quirk and his colleagues, whose inception was in 1959.

so-called *hyōgen ito* “expressive intent”), (c) syntax of spoken language, and (d) intonation. The last chapters of each volume are devoted to the proposal of “generalized sentence-types” (*sōgōteki bunkei*), which are proposed on the basis of the synthesis of (a)–(d) above. Here, only the result of the intonation study will be introduced. The principal investigator of the intonation study was Norio Yoshizawa.

Compared to the results reported in NLRI (1955), there is one important piece of progress in the study of intonation in NLRI (1960, 1963): the distinction of five basic intonation types in sentence-final intonation. They were recognized on the basis of subjective descriptions of corpus data, but a preliminary F0 analysis by means of an analogue pitch recorder was also utilized to provide supplementary experimental evidence.

The five types included flat (*heichō*), falling (*kōchō*), type 1 rising (*shōchō 1*), type 2 rising (*shōchō 2*), and, the special class named type-@ (@ *gata rui*). This was a very early inception of the phonological classification of phrase-final intonation which is known today as boundary pitch movements, or BPM.

The distinction of five sentence-final intonation types introduced in this study set off at least two important debates on Japanese intonation. First, there is a long-standing debate about the necessity of the distinction between the so-called “falling” and “flat” types for the characterization of sentence-final rendition. See Kori (2008) for details of the debate. Part of the debate overlaps with the American debate about the domain of so-called final lowering in Japanese (Poser 1984; Pierrehumbert and Beckman 1988). See also section 3.5 below.

Second, the discrimination of two different rising intonation types was refined later in Shin Kawakami’s seminal work (Kawakami 1963) that classifies type 1 rising intonation into three subclasses, namely, normal rise (*futsū no jōshōchō*), floating rise (*ukiagari chō*), and, incredulity question (*hanmon no jōshōchō*).[2] In addition, Kawakami renamed the NLRI’s type 2 rising intonation as insisting rise (*tsuyome no jōshōchō*). Currently, this four-way classification of rising intonation is widely accepted among Japanese phoneticians.

Lastly, there is a spin-off study of NLRI (1960, 1963) to be noted here, viz., Oishi (1959). In this paper, Oishi gave an overview of various possibilities of realizing “prominence”, or focus if we adopt today’s standard terminology, in the phonetics of Tokyo Japanese, and pointed out the presence of a hitherto unknown realization type of prominence that he called *atodaka gata*, in which the penultimate mora of a word carries local non-accentual prominence. A fuller understanding of the nature of this special prominence, known currently as the “PNLP”, was brought about in 2011 by the analysis of CSJ. See section 3.5 below.

The age of pre-computerized speech corpora came to an end after the publication of NLRI (1963). For some unknown reason, the NLRI abandoned the line of research projects on spoken language that it developed in the 1950s and 1960s as

2 English translation of Kawakami’s intonation types is by Maeda and Venditti (1998).

a world pioneer. Because NLRI was the only institution that had the potential of conducting such costly studies, corpus-based linguistic study of spoken language entered a long period of stagnation until the end of the 1990s. It was speech engineering studies that filled the long gap caused by the stagnation in linguistic studies.

3.2 Analysis of segment duration

As is well known, speech science in Japan has a long history of contributions that goes back to Chiba and Kajiyama's contribution to the basic understanding of the speech production mechanism in 1942.

Post-war, Japanese researchers continued to make substantial contributions to the science and processing of spoken language. The study of corpus-based speech synthesis is one such contribution.

Text-to-speech synthesis, i.e., the reading aloud of a given text by computer, requires, among many other things, precise control of segment durations. In phonetics, and phonology, Japanese has long been treated as a typical mora-timed language. The linguistic principle of mora-based isochronism per se, however, turned out to be insufficient for the synthesis of natural-sounding speech.

If all moras of a synthesized speech have exactly the same duration, the resulting speech sounds awkward. Factors like the following are known to have significant influence on the systematic variation of segmental duration in Japanese: (a) inherent duration of the target phoneme, (b) the phoneme's compressibility in terms of duration, (c) temporal compensations between adjacent phonemes, (d) mora timing, (e) lengthening of content words and shortening of functional words, (f) lengthening at the end of a phrase and shortening at the beginning of a phrase, (g) shortening due to the increase in the number of moras in a phrase, and (h) overall speaking rate. (Sagisaka 1993).

These factors were discovered by statistical analyses of read speech data of carefully prepared sentences (Sagisaka and Tohkura 1980, among others), which can be regarded as an early stage of corpus-based speech analysis. The weightings of these factors (i.e., control parameters of text-to-speech conversion) were estimated by use of various techniques of regression analysis.

The performance of the statistical computation model of segment duration was much better than those of the traditional rule-based models. But the new model had its own problem. It could be very unstable under certain circumstances due mainly to the limitation of the read speech data from which the control parameters were acquired.

Hence arose the necessity of speech corpora much larger in size and more authentic with respect to the distribution of phonetic/linguistic factors (phoneme balance is just one example of such factors). Representative studies along this line include Takeda, Sagisaka, and Kuwabara (1989), Kaiki, Takeda, and Sagisaka (1991)

and Iwahashi and Sagisaka (2000), among many others. These studies revealed convincingly that the computation of segment duration is not as simple a phenomenon as supposed by many phoneticians.

It is worth noting here that the vital importance of hierarchical prosodic structure was shown for Japanese for the first time by these speech processing oriented studies using corpora. The above-mentioned studies have direct relevance to the understanding of many basic issues in phonetics like mora- and syllable-timing, final lengthening, segment reductions, and so forth. See also Warner and Arai (2001), who cast doubt on the effectiveness of a mora-timing hypothesis based upon the analysis of spontaneous speech (see also Otake, this volume, for full discussion of moras and mora-timing).

The application domain of the technique of corpus-based acquisition of control parameters is not limited to the control of durational features. In fact, the technique was applied to the acquisition of control parameters for speech fundamental frequency (F0) and speech power (Hirai et al. 1995, for example). From a point of view of linguistics and phonetics, however, it is the studies of durational control that are the most interesting.

3.3 Analysis of speaking style and phonetic variations

In the studies reported in the previous subsection, control parameters were estimated by analysis of read speech corpora. There are, however, good reasons to believe that the best contribution of corpus-based analysis can be found in the analysis of spontaneous speech. Study of speaking style is a typical example.

There seems to be a wide consensus among researchers of phonetics, sociolinguists, and speech engineers that the way we speak differs considerably depending on the social settings in which we speak, but it is only recently that the comparison of read and spontaneous speech has become possible.

Nakamura, Iwano, and Furui (2008) compared the spectral characteristics of the spontaneous speech in CSJ and the read speech of JNAS corpus and found a systematic correlation between speech spontaneity and the shrinkage of spectral space. Spectral space (articulatory space) shrinks as the speaking style becomes more spontaneous (i.e., more casual). They also showed that the reduction in the spectral space triggered lowering in phoneme recognition accuracy.

In this study, spectral characteristics of segments were represented by the *Mel-Frequency Cepstrum Coefficients* (MFCC) instead of the formant parameters (central frequency and bandwidth) which are widely used in the study of phonetics. MFCC is a common way of representing spectral information in ASR. MFCC is superior to formant in that it can be computed without relying upon complex algorithms with the consequence that MFCC is much more robust (hence reliable) than formants. Another characteristic of MFCC is that its computation is much less influenced by F0 than in the case of formants.

It is not only the articulation (i.e., the spectral space) but also the linguistic specification of words that is under change in casual speech. For example, the adverb *yahari* 'after all' has at least five different word forms other than /jahari/, viz., /jappari/, /jahasi/, /jappasi/, /jappa/, and /pasi/ according to the CSJ. Among them, the percentages of authentic /jahari/ and casual /jappari/ are 61.5% and 30.3% respectively in APS, but 29.9% and 52.1%, respectively, in SPS.

As in this example, the word-form actually used can be different from its "proper" form (which is usually registered in dictionaries) depending on the speaking style. Maekawa (2009) analyzed the CSJ and showed a list of words that tend to be realized in non-"proper" word forms. The list includes variations like /iu/~/ju:/, /no/~/n/, /nani/~/nan/, /nihon/~/nippon/, /jahari/~/jappari/~/jappa/~/jahasi/, etc. He also showed that on average more than 95% of variable tokens could be covered by knowing the three highest frequency variants of all variable words. The drawback of this study is that the number of words whose variations can be analyzed in this way is very limited even when a corpus like CSJ is utilized. It covers only a few dozen words.

Akita and Kawahara (2005) proposed a generalized model that could predict the occurrence probabilities of various non-"proper" word-forms starting from the input of the "proper" word-form. The model, which is capable of treating 256 different variation patterns, was learned statistically from the variations recorded in the CSJ. The resulting 'pronunciation dictionary' turned out to be effective in improving the performance of ASR systems for spontaneous speech.

Unlike all the above-mentioned studies that evaluated segmental features, Maekawa (2011a) evaluated the differences of prosodic features due to the difference of speech registers in the CSJ (namely, the differences among APS, SPS, dialogue, and reproduction). It was shown that it was possible to classify correctly the speech registers of the talks in the CSJ-Core with 85% accuracy (closed data), given the frequency information of the X-JToBI labels and the speaking rates.

The last two studies aimed at holistic evaluation of phonetic variation in general. Needless to say, this kind of study was impossible without a reliable corpus of spontaneous speech, hence they were absent before the release of the CSJ.

On the other hand, there are 'traditional' variation studies that analyzed individual phonetic variations in some depth. Maekawa and Kikuchi (2005) analyzed devoiced vowels in the CSJ and concluded that vowel devoicing was not as predictable as hitherto believed (see Fujimoto, this volume, for full discussion of vowel devoicing). Frequently, there are cases like devoicing of non-high vowels (/e/, /o/, and /a/), devoicing of high vowels in front of voiced segments, and non-devoicing of high vowels in the typical environment of devoicing, i.e., the case when high vowels are both preceded and followed by voiceless consonants. On the basis of these and other observations, they proposed a hypothesis that the probability of high vowel devoicing in the typical devoicing environment is conditioned by the spectral distance between the two voiceless consonants on both sides of the high vowel in

question, especially in the case where more than two consecutive moras are in the typical environment for devoicing.

Several years later, Kawatsu and Maekawa (2009) conducted acoustical analysis of the CSJ and presented evidence supporting the hypothesis that there was a positive correlation between the spectral distance between consonants and the probability of vowel devoicing.

These two studies challenged the traditional view that devoiced high vowels were conditioned variants of their voiced counterparts and showed that devoicing was a probabilistic event. The same kind of challenge could be cast on other phonetic variations. Maekawa (2010a) analyzed the variation in the manner of articulation of the /z/ consonant (between fricative [z] and affricate [dz]), and concluded that the variation was not a conditioned variation. Traditionally, the variation has long been described as a varying as a function of the location of the consonant in a word, viz., affricate in word-initial position, and fricative elsewhere. In the analysis of the CSJ-Core the mean probability of affricate realization showed a nearly monotonic relationship with a newly introduced phonetic parameter called TACA (Time Allotted for Consonant Articulation) that approximated the total amount of time that a speaker could use for the articulation of the consonant. The mean prediction accuracy of the manner of articulation by TACA is higher than 74%, and when coupled with other variables, the accuracy reaches a level of about 80%.

Maekawa (2010b) showed that TACA was also effective in the analysis of the weakening of stop articulation in Japanese voiced stops, /b, d, g/. The probabilities of the weakening of [b], [d], and [g] into [β], [ð], and [ɣ] are inversely correlated with the TACA values.

3.4 Analysis of filled pauses

Spontaneous speech is colored by various sorts of speech disfluency. By speech disfluency is meant phenomena like filled pauses (or simply 'fillers'), false starts, self-repair, fragmented words, and so forth. Although most linguists tend to regard these speech phenomena as peripheral to linguistic systems, they play important cognitive roles in speech communication. In this section, only the studies regarding filled pauses will be presented.

Watanabe et al. (2008) and Watanabe (2010), for example, analyzed the CSJ and showed that utterances tended to be longer or more complex when they were immediately preceded by filled pauses than when they were not. They also showed by experiments that the existence of filled pauses caused listeners to expect that the speaker was going to refer to something that was likely to be expressed by a relatively long or complex constituent. Watanabe (2013) examined the same hypothesis by using the distance of dependency relationship as a measure of the complexity of the upcoming phrases, and arrived at the same conclusion as in her earlier studies.

Similarly, Den (2009) analyzed phonetic prolongation of what he called 'clause-initial mono-words' (words like /de/ and /ma/ occurring at the beginning of a clause), and found that, in the CSJ, the prolongation rate was higher in SPS than in APS, due probably to the difference of cognitive load of speech planning: speakers have higher cognitive loads in SPS because SPS is generally more spontaneous than APS.

By the way, as can be readily understood by comparison between Japanese and English, filled pauses are language-dependent at least to some extent, but exploration of the place of filled pauses in linguistic structures has been left almost untouched in current linguistic studies.

Maekawa (2013b) examined the question of how F0 values of filled pauses were determined in speech, and concluded that it was possible, at least to a certain extent, to predict the F0 values of filled pauses given the F0 values of the boundary tones of the immediately adjacent accentual phrases. This study suggests that filled pauses have no specification of phonological tones; this is contrary to the assumption held in the design stage of the X-JToBI annotation scheme in which a tone was specified for each filled pause (see Maekawa et al. 2002).

There are also application-oriented studies of filled pauses. Ohta, Tsuchiya, and Nakagawa (2007) proposed a prediction model of filled pauses based upon analysis of the CSJ. The model was used to construct a new language-model including filled pauses starting from corpus data that does not include filled pauses; this is a technique called domain-adaptation of language models. Domain adaptation is necessary because corpora containing filled pauses are quite limited in number compared to corpora without filled pauses (corpora of written language, for example) and the inclusion of filled pauses is badly needed for the processing of spontaneously spoken language.

Ohta, Tsuchiya, and Nakagawa (2010) also proposed a prediction model of silent pauses (as opposed to filled pauses) whose occurrence location does not match any kind of syntactic boundary (word-internal silent pause, for example). Needless to say, this kind of silent pause can be regarded as a sort of speech disfluency. In this respect, Kagomiya et al. (2007) reported in their study of impression rating score (see section 3.7 below) that the occurrence rates of silent pauses and word fragments had a strong influence on the impressions that listeners receive from various speech in the CSJ. See section 3.7 below for more details.

Lastly, Ishihara (2010) examined the usefulness of filled pauses as a factor of speaker discrimination, and found gender differences in filled pauses in the CSJ. Female speakers tend to use more variable filled pauses than male speakers.

To sum up, the study of speech disfluency is inseparable from the analysis of spontaneous speech corpora. In Japanese, as well as in many other languages, it is one of the fields of phonetic research that has benefited most from spontaneous speech corpora. See Maruyama (2013) for speech disfluencies other than filled pauses in Japanese.

3.5 Analysis of dialogue and discourse

Like the study of filled pauses, study of dialogue is one of the fields that has benefited greatly from speech corpora. Dialogue speech has several characteristics that are completely lacking in monologue speech.

Turn-taking of speakers is one such characteristic. Koiso et al. (1998) analyzed syntactic (part of speech) and prosodic (duration, F0 contour, peak F0, peak energy, etc.) features at the point of turn-taking by speakers occurring in the *Chiba University Japanese Map Task Dialogue Corpus* with the view of clarifying the relative importance of these features for the smooth realization of turn-taking and back channeling in spontaneous dialogue. Their conclusion included, among other things, that, in general, syntax had a stronger contribution than any single prosodic feature, but that prosody as a whole contributed as strongly as, or even more strongly than syntax.

Using the same corpus, Ohsuga et al. (2006) conducted an experiment of predicting turn-taking using only the prosodic features of speech. The results of their decision-tree analysis achieved a higher than 80% correct discrimination rate (open data). Similarly, Enomoto (2007) examined the concomitant effect of argument structure (as expressed by the presence of case particles) and what she called "utterance final elements" (as represented by auxiliaries – *desu* and *masu* – and phrase final particle – *ne* and *yo* –) on the latency of turn-taking. She found that the latency was considerably shorter when there were utterance final elements.

Lastly, Koiso and Den (2010) is a pilot trial to expand the prediction model to cases involving overlapping speech, i.e., the case where two speakers speak simultaneously at the point of turn-taking.

The acoustic features analyzed in the above studies are all local features in that they are the properties of segments, moras or, at the maximum, syllables. There are however, global prosodic features that expand over phrases. Koise, Shimojima, and Katagiri (1998) analyzes the relationship between the change in speaking rate and the information structure of dialogues. They found correlations between deceleration in speech and the opening of new topics in dialogues, and conversely, acceleration in speech and the absence of new topic opening.

In dialogue, speakers tend to accommodate each other with respect to the way they speak. Nishimura, Kitaoka, and Nakagawa (2009) compared F0 contours of two participants of CSJ dialogues and found weak positive correlations. The correlation was especially clear in highly "lively" dialogues. See also section 3.7 below for this study.

Dialogue speech was also studied in the field of speech processing with a view to finding basic design criteria for a natural language man-machine dialogue system. Data used in this area were acquired mostly using the "Wizard of Oz" protocol wherein an experimenter (the "wizard") simulates the behavior of an intelligent computer application under investigation.

Although some of the dialogue data obtained by this protocol was publicly available (Itou et al. 1999), the corpus has not been widely used, most probably because the characteristics of the collected dialogues were strongly restricted by the purpose of the simulated system.

There are also corpus-based studies about the phonetic cues of discourse boundary and structure. Fon (2002) conducted a cross-linguistic comparison of durational features as cues for discourse boundary markers using elicited spontaneous speech in English, Japanese, and two dialects of Mandarin (*Guoyu* and *Putonghua*); she concluded that final lengthening of boundary syllable and syllable onset interval were the most universal cues of signaling structural boundaries.

Yoneyama, Koiso, and Fon (2003) analyzed the relationship between the prosodic labels of the X-JToBI annotation and the discourse boundary using a subset of the CSJ-Core. All prosodic features examined – declination and resetting of accentual F0 peaks around a discourse boundary, choice of BI (boundary indices) at the boundary, and the choice of boundary tones at the boundary – showed significant correlation with the strength of the discourse boundary.

As suggested by this study, choice of boundary tones, especially the use of complex boundary tones or boundary pitch movements (BPM) can be utilized as a cue to discourse boundaries. Maekawa (2011b) examined the distribution of a special BPM called 'PNLP' (penultimate non-lexical prominence) in the CSJ-Core in relation to various utterance boundaries, and found that PNLP appears most typically shortly before the end of deep utterance boundaries that correspond, presumably, to the end of a discourse topic. See the study of Oishi summarized in section 3.1 above.

Generally speaking, a mutual relationship between the syntactic and/or discourse boundaries and the prosodic characteristics near the boundaries is one of the most interesting issues in the phonetic study of discourse structure, and many studies are currently underway. For example, Koiso and Den (2013) reported a correlation between the probability of BPM and the complexity of upcoming phrases (see the studies by Watanabe mentioned in section 3.3 above).

Maekawa (2013c) examined final lowering in the CSJ and found that final lowering is observed in all clause boundary classes, and, further, the degree of lowering is positively correlated with the strength of syntactic boundaries. It seems that final lowering is not a mere signal of the end of an utterance; it rather signals various degrees in the depth of syntactic and/or discourse boundaries.

Lastly, there are a few pilot studies that analyzed the relationship between topic structure and prosody. Nakagawa, Asao, and Nagaya (2008) and Nakagawa, Yokomori, and Asao (2010) examined respectively the relation between the prosodic phrasing of syntactically right-dislocated phrases and their information structure, and, the function of intonation units within discourse structure and information structure.

3.6 Analysis of infant speech and infant-directed speech

The study of language acquisition often provides vital evidence for both psychological and linguistic studies of language (see Ota, this volume, and Hirata, this volume, for full discussion of L1 and L2 acquisition). At the same time, there is consensus among researchers that studies of infant speech and/or infant-directed speech (IDS) are impossible without corpora of naturalistic speech data.

One of the basic issues in this area is the determination of the period when infants start to use vocal signals as a manifestation of language. Analysis of longitudinal corpora provides basic information regarding this fundamental question.

Amano, Nakatani, and Kondo (2006) examined developmental changes in the F0 production of infants and their parents' IDS using the *NTT Infant Speech Database*. They found that both infants and parents showed critical changes in their F0 in terms of within- and between-utterance variability at the time when the infants started to use two-word utterances. They suggest that these changes are the reflection of the beginning of communication by means of "language".

There is a similar study of segmental sounds. Ishizuka et al. (2007) examined the development of vowel space of infants of 4 to 60 months in order to determine when and how infants become able to produce categorically distinct vowels. It turned out, also by the analysis of the NTT database, that infants started producing categorically separate vowels by 24 months. Interestingly, the critical periods reported by these studies do not coincide. The onsets of two-word utterances observed in Amano, Nakatani, and Kondo (2006) were found at around 18 months.

Another important issue in this field is the clarification of the mechanism by which infants acquire phonemic category contrasts. In this respect, analysis of IDS plays a crucial role. Bion et al. (2013) examined the case of phonemic vowel quantity (length) in Japanese using the *RIKEN Japanese Mother-Infant Conversation Corpus* (Mazuka, Igarashi, and Nishikawa 2006). They found that it was almost impossible to acquire the contrast between short and long vowels simply based upon the distributional difference between their acoustic durations. The conjoint distribution of short and long vowels forms a unimodal distribution due mainly to the considerable difference in the number of tokens between the two categories (more than 90% of vowels are short in the corpus). Based mainly upon this finding, the authors suggested that the learning of phonemic contrast was helped by the simultaneous learning of the lexicon, especially by the presence of minimal pairs.

The possibility of simultaneous distribution-lexicon learning is examined by means of simulation in Martin, Peperkamp, and Dupoux (2012). This paper proposed a new algorithm by which infants acquire the phoneme inventory of a language in the first year of their lives. The new algorithm makes use of not only distributional properties of the allophones and the phonetic similarities among them, but also the information about what they call a "proto-lexicon," an approximation of the language's lexicon estimated by means of highly frequent n-grams. They applied the

new algorithm to the corpus data of Japanese (CSJ) and Dutch and obtained promising results.

Lastly, Igarashi et al. (2013), who examined intonational exaggeration in the IDS of RIKEN corpus, points out the importance of a phonological theory of intonation in the study of IDS. They showed that in Japanese IDS, intonational exaggeration (pitch range expansion) is observed almost exclusively in the BPM. The important point is that the exaggeration can't be recognized if the pitch range of the utterance as a whole is analyzed, because the part of utterance preceding the BPM behaves differently from the BPM, thereby canceling the exaggeration in the BPM.

3.7 Analysis of paralinguistic and extra-linguistic information

As is well known, speech does not convey linguistic information alone. It also conveys so-called 'paralinguistic' and 'extra-linguistic' information. Studies in this field have traditionally been conducted on using experimental data. Recently, however, corpus-based studies have become available. As a matter of fact, Igarashi et al. (2013), mentioned in the preceding section, belongs to this type. We will start by summarizing the studies of extra-linguistic information.

Wada, Shinozaki, and Furui (2010) tried to estimate the age of the CSJ speakers by the combinations of various spectrum-based features used in ASR studies and the statistical classification techniques including SVM (support vector machines) and SVR (continuous support vector regression). Omitting the details, the results showed that it was possible to estimate the speakers' age with the absolute error of 7.3–10.9 years. Note that in CSJ, the data about speakers' age is given in five year increments.

Maekawa (2012) tried to predict the age group of the CSJ speakers using the data of X-JToBI label frequency prepared in Maekawa (2011a). When the speakers were classified into four age groups, the correct classification rate of linear discriminant analysis was about 50%. The same study also tried prediction of speaker gender, and the correct classification rate was about 80%.

Kinoshita, Ishihara, and Rose (2009) examined the usefulness of the F0 distribution information for the sake of forensic phonetics (namely, speaker recognition). Their study remains in the preliminary stage, but the reported results sound promising.

There is wide disagreement in what is meant by the word 'paralinguistic feature' or 'paralinguistic information', but the number of corpora annotated with respect to "paralinguistic" characteristics is very limited in any sense of the word. CSJ is annotated with respect to so-called *impression rating score* or IRS. IRS is a subjective evaluation of various impressions that speakers receive from a talk. It covers various impressions like spontaneity, formality (as opposed to casualness), skillfulness, speediness, and so forth.

The IRS annotation of the CSJ-Core was conducted based on a psychological scale newly constructed for the monologue spontaneous speech in the corpus.

Yamazumi et al. (2005) provides detailed information about the construction of the psychological scale, and Kagomiya et al. (2007) provides information how the psychological scale was utilized in the task of corpus annotation.

The last half of Kagomiya et al. (2007) is devoted to pilot analyses of the IRS data. They found out, for example, that the "skillfulness" and "activity" perceived by listeners of a talk were both strongly correlated with the occurrence rate of silent pauses. In another paper (Kagomiya et al. 2008), the same authors analyzed the subjectively perceived speaking rate, and found out that perception of "slowness" is more influenced by the occurrence rate of silent pauses than the physical speaking rate.

Similarly, the last half of Nishimura, Kitaoka, and Nakagawa (2009) reports the results of their own IRS evaluation on the "liveliness," "familiarity" and "informality" of the CSJ dialogue data and concluded that evaluated liveliness correlated more closely with F0 than speech power.

As mentioned previously in section 2.2, the UUDB is annotated with respect to the perceived emotional states of the speakers. Mori et al. (2011) conducted acoustic analyses of the corpus and reported the importance of voice-quality parameters (to be more concrete, the ratio between the periodic and aperiodic component in the F0) for the recognition of the paralinguistic information (speakers' emotional status). Importance of voice quality in dialogue speech was also stated in Ishi, Ishiguro, and Hagita (2010).

4 Recapitulations and prospects

4.1 Trends and problems of Japanese speech corpora

As is clear from the overviews presented so far, not all speech corpora mentioned in section 2 are utilized with similar frequencies. CSJ is by far the most frequently utilized corpus. There seem to be multiple reasons for this. CSJ is large in size and rich in annotation. Moreover, it covers multiple speech registers, i.e., monologue, dialogue and read speech. These features match exactly the tendencies of corpus development in recent years, which can be summarized by four simple phrases: from small to large, from read to spontaneous, from specific to balanced, and from linguistic to paralinguistic.

There are, however, important speech registers that are not covered by the CSJ or any other corpora. Those registers include task-free (as opposed to task-oriented) conversation and multi-party conversation, among many other registers. The former register is needed, among other things, to explore the variability of speech at the lower end of speech formality.

On the other hand, the latter register is needed to explore the mechanisms by which complex human communication is conducted in the real world. Needless

to say, a pure speech corpus is not sufficient for a full understanding of the complexities of human interaction. Visual information like eye-contact among speakers, various body gestures, and their complex interactions may play important roles in communication in the real world. Hence there arises the necessity of multi-modal corpora in this field. See the relevant chapters in Bono and Takanashi (2009) that summarize the techniques of multi-party corpus annotation and corpus analyses.

Further refinement of corpus annotation criteria currently in use is also needed. For example, Den et al. (2010) propose a new annotation scheme of conversation data that may replace various annotation schemes utilized in the CSJ.

There are still three more types of speech corpora whose development is badly needed for Japanese phonetics and phonology. One of them is a corpus that collects systematically the manifestations of paralinguistic and extra-linguistic features of speech. The recently published *Online Gaming Voice Chat Corpus* (Arimoto et al. 2012) is one such corpus. Also, the ongoing *SEN* project led by Hideaki Kikuchi (Miyajima, Kikuchi, and Shirai 2011; Kikuchi, Miyazima, and Shen 2013) is of particular interest. Although this is a corpus of "acted" paralinguistic features, the specification of paralinguistic information in this corpus is highly systematic and covers a very wide area of paralinguistic information that no other corpora have tried to cover.

The second type is a corpus of dialect speech. At the current time, no corpus of Japanese dialects is available, but a pilot project is currently underway at the Department of Language Change and Variation of NINJAL (National Institute for Japanese Language and Linguistics) under the leadership of Nobuko Kibe. The aim of the project is to construct a dialect corpus by compiling pre-existing recordings of dialect speech that were collected in the early 1980s in a collaboration between the Agency of Cultural Affairs (*Bunkachō*) and the former NLRI.

The third type is a corpus that collects audio recordings of Japanese in past eras (after the invention of audio recording technique, needless to say). In other words, a corpus of movies and radio- and TV-programs. No such corpus seems to have yet been compiled for Japanese (or any other language). The existence of such a corpus would enable direct observation of phonetic changes (as opposed to observation of changes in apparent time) that have occurred in the past seventy years or so.

In this respect, Sato (2011) seems to be an interesting precursor in this field. He analyzed the distribution of shot sequences and silent pauses in the movies of Yasujiro Ozu and found that the shot sequences were carefully edited so that the audience perceives the rhythm of conversation that is specific to Ozu's movies.

Before closing this subsection, the importance of resource sharing is to be stressed. As pointed out in Maekawa (2013a) and other studies, one of the most important premises of modern corpora is that they will be publicly available. A corpus needs to be a shared resource among the researchers of related fields.

In this respect, it is unfortunate that some large-scale corpora developed in recent years are not open for public use. RIKEN's infant corpus mentioned repeatedly

in section 3.6 is an example. Public release of an important corpus like this would be a great benefit for the researchers of the related fields. Another example is the corpus built by the CREST-ESP project (Campbell 2000; Douglas-Cowie et al. 2002).

A corpus built with public funding should become publicly available as soon as the first-hand analyses by the members of the construction group are finished, unless there are special reasons not to do so. In the cases of the above-mentioned corpora, it is widely believed that it has mainly been the issue of copyright clearance that has prevented them from being publicly available. Corpus constructors should make every effort to exclude such 'special' reasons at the time of corpus design and during the course of corpus construction.

4.2 Trends and problems of corpus-based analyses

In the preceding subsection, four tendencies in recent corpus development were presented. We can summarize the recent tendency of corpus analysis by incorporating the four tendencies into a single phrase: 'analysis of *spontaneous* speech by means of *large*-scale *balanced* corpus with linguistic and *paralinguistic* annotations'.

The fields of language research that fit most readily to this phrase include the study of socio-phonetic variation, the study of speech development, and the study of dialogue (or conversation) speech. These fields will certainly continue to depend heavily upon corpus analysis in the coming decades, probably with further corpus enhancement along the lines suggested in the previous subsection.

The investigation of the phonetics and phonology of paralinguistic information is still in its preliminary stage, but it will develop quickly if data like the SEN-corpus (see section 4.1 above) become publicly available. It is of particular interest to examine the variation of different phonological entities (segmental phoneme and tones) under the influence of paralinguistic information, which has been examined so far only in experimental settings (Maekawa 2004a).

On the other hand, there are fields of speech research that have almost no connection with corpus studies at the present time, despite evidence suggesting the possibility of a connection. One such field is the study of phonology as opposed to phonetics.

For example, it seems that a basic assumption of Optimality Theory in phonology (gradient wellformedness) is fully compatible with data obtained through analysis of either the spontaneous speech corpora or the infant speech corpora, which include many 'incorrect' word forms. As far as Japanese is concerned, however, it seems that no full-fledged study has been conducted on this theme.

Another field of possible connection is the study of historical changes. In the study of phonological changes, the phonetic process of the changes and the sociolinguistic background of the changes are often discussed. So far, however, the studies have been based upon conjecture rather than empirical analyses. There is a

chance that the process of historical change can be simulated by the analysis of actual speech data in the corpus.

The phonetic processes behind changes like weakening or enhancement of articulatory gestures or gradual shifts in the place of articulation under the influence of various prosodic conditions all exist in present-day spontaneous Japanese; fine analyses of the processes may provide important information about changes that occurred in the past.

Similarly, use of large-scale corpora enables, for the first time in the history of linguistics, reliable measurement of the so-called function load of phonemes and their change due to sociolinguistic factors such as speech registers. Here again, studies on these themes have not been conducted, with the exception of Oh et al. (2013), which tried to compare the functional loads of various phonemes in different languages, including Japanese.

Lastly, it is curious that virtually no phonetic analysis has been conducted in the field of JSL (Japanese as a Second Language) studies even though there are corpora of JSL learners with audio signals (see Table 1). Probably, it is a lack of knowledge of speech analysis and the techniques of corpus query on the part of JSL researchers on the one hand , and a lack of appropriate phonetic annotation on the other, that interfere with the development of phonetic analyses in the field of JSL study.

So far, problems have been pointed out in terms of research fields. There are also problems in research methods. A problem of corpus analysis that researchers of these fields encounter quite often in their daily practice is the application of statistical methods. Traditional methods of statistics (i.e., various inferential statistical tests including t-test, ANOVA, etc.) in corpus analysis can be problematic when the data supplied by a corpus is too massive in size. As is well known, the standard deviation of the sample mean is given by $1/\sqrt{N}$, where N is the number of samples randomly selected from the population. This is a consequence of the so-called "law of large numbers" or "central limiting theorem."

Suppose a case where two means are compared to find if there is statistically significant difference, and each of the mean values is computed based upon 10,000 samples from a corpus. In this case, the standard deviations of the sample means become 1/100. The consequence is that a very small difference between the two sample means – a difference that can hardly have any meaningful effect based on empirical knowledge – becomes significant. Traditional setting of the significance-level (0.05 or 0.01) doesn't make much sense in a case like this.

Some researchers adopt new a significance-level (0.0001 for example), but this cannot be a complete solution. It is nearly impossible to establish a new significance-level that fits all corpus analyses, because there is no theoretical limit in the number of "excessive" samples, and the number of samples obtained from a corpus differs considerably from one analysis to another (from ten to one thousand, for example) depending on the occurrence frequencies of the linguistic variables.

A fundamental solution to this problem is to shift from traditional statistics to so-called Bayesian statistics that does not rely upon the application of the central limiting theorem. Unfortunately, however, such a shift is not an easy task. Application of Bayesian statistics requires some specialized knowledge in statistics and certain computing skills. In this respect, innovation in graduate-school education is badly needed in fields like linguistics and phonetics.

Another problem of corpus analysis can be found in the way researchers query speech corpora. Unlike the written language corpora, it is often difficult to query the annotation information provided with a spoken language corpus, because annotation of phonetic events like the X-JToBI is inherently multidimensional.

Researchers often find it necessary to make complex queries encompassing both linguistic and prosodic structures. For example, Maekawa (2010a) extracted for all /z/ consonants in the CSJ-Core more than 30 information items including (a) the duration of /z/, (b) the durations of the four preceding phonetic segments, (c) the durations of the three following phonetic segments, (d) the POS (part-of-speech) of the SUW (short-unit-word) in which the /z/ appeared, (e) the POS of the LUW (long-unit-word) in which the /z/ appeared, (f) the location of the mora including the /z/ (the mora, hereafter) in an accentual phrase, (g) the location of the mora in the LUW, (h) the location of the mora in the SUW, (i) the lemma of the SUW, the lemma of the LUW, (j) accentedness of the mora, (k) accentedness of the preceding SUW, (l) BPM, if any, at the end of the preceding accentual phrase, (m) the break index of the preceding SUW, (n) the word type (*goshu*) of the SUW, (o) the speaking rate of the accentual phrase, and so forth. Moreover, meta-information like the gender and age of speakers, and type of talk (APS, SPS etc.) were also used in the analyses.

Extraction of this information (which has to be repeated frequently as the analysis proceeds) from the corpus is not an easy task for average linguists and phoneticians. Kikuchi and Maekawa (2007) used query scripts written in the XSLT language to extract the information from the XML formatted data of the CSJ-Core at the time they wrote the paper, but the process was later replaced by the use of the RDB (relational database) version of the CSJ-Core (http://www.ninjal.ac.jp/corpus_center/csj/data/rdb-outline/).

Writing the queries for the RDB version by means of SQL language is much easier compared to programming in XSLT, and the query time is drastically shorter than that required for XSLT query of XML documents. As shown by this example, dissemination of corpus analysis cannot be separated from the implementation of the corpus data.

Finally, it is necessary to point out the need for a link between experimental and corpus-based speech studies. Despite the belief of fanatic proponents of corpus studies, experimental- and corpus-based approaches are not incompatible. Rather, in most cases, they are in a complementary relationship.

Some of the corpus-based studies reported in section 3 proposed new hypotheses about speech production based exclusively upon the analyses of spontaneous

speech corpora (the mechanism of variation of /z/, for example); the validity of the hypotheses will become all the more clear if the patterns found in spontaneous corpora can be reproduced in experimental settings.

Although it can be very difficult to design such experiments for some phonetic events (analysis of PNLP, for example), the reproduction experiments are worth conducting. Any disagreement between the experimental and corpus-analyses strongly suggests incompleteness of the analyses; the incompleteness may be found in either the experimental study or corpus-based study, or perhaps in both.

4.3 Future prospects

Although not widely known, corpus-based analysis of spoken language was born in Japan as early as in 1955. It was only ten years after the end of the Pacific War and the complete destruction thereby of the infrastructure of the country.

The birth was supported by the young researchers of the NLRI who were mostly trained in traditional Japanese linguistics (*kokugogaku*). One can feel clearly their eagerness to open up new perspectives of language study in the post-war society when one reads their reports (NLRI 1955, 1960, 1963). However, the field soon entered a long period of stagnation that covers the 1970s and 1980s.

A period was put to the stagnation in the beginning of the 21st century by the release of the CSJ and other large-scale corpora of spontaneous Japanese. Corpus-based study of spontaneous Japanese is now reviving quickly. This time, however, the revival is supported mainly by the people working outside the field of linguistics and phonetics.

They are mostly people working in the fields of speech processing and psycholinguists who are interested in the mechanisms of complex human interaction including, but not limited to, language. Language development researchers also contribute greatly to the revival.

Perhaps we are now witnessing the second heyday of corpus-based phonetics in this country. To sustain this cherished moment, it is of prime importance to maintain the interdisciplinary and transdisciplinary nature of the field, thereby expanding continuously the application domains of the outcomes from the field. Lack of such effort was probably the main reason for the quick decay of corpus-based phonetics in the past.

References

Akita, Yuya and Tatsuya Kawahara. 2005. Hanashi kotoba onsei ninshiki no tame no han'yōteki na tōkeiteki hatsuon hendō moderu [Generalized statistical modeling of pronunciation variations for spontaneous speech recognition]. *The Transactions of the Institute of Electronics, Information and Communication Engineers* D-II, J88-D-II, 9. 1780–1789.

Amano, Shigeaki, Tadahisa Kondo, Kazumi Kato and Tomohiro Nakatani. 2009. Development of Japanese infant speech database from longitudinal recordings. *Speech Communication* 51. 510–520.

Amano, Shigeaki, Tomohiro Nakatani and Tadahisa Kondo. 2006. Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *Journal of the Acoustical Society of America* 119(3). 1636–1647.

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Bwyneth Doherty, Simon Garrod, Stephen Isard, Jacquline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson and Regina Weinert. 1991. The HCRC map task corpus. *Language and Speec*h 34(4). 351–366.

Arimoto, Yoshiko, Hiromi Kawatsu, Sumio Ono and Hitoshi Iida. 2012. Naturalistic emotional speech collection paradigm with online game and its psychological and acoustic assessment. *Acoustical Science and Technology* 33(6). 359–369.

Bion, Ricardo, A. H., Kouki Miyazaki, Hideaki Kikuchi and Reiko Mazuka. 2013. Learning phonemic vowel length from naturalistic recordings of Japanese infant-directed speech. *PLoS ONE* 8(2): e51594. (doi:10.1371/journal.pone.0051594).

Bono, Mayumi and Katsuya Takanashi (eds.). 2009. *Taninzū intarakushon no bunseki shuhō* [Analysis method of multi-party interaction]. Tokyo: Ohmsha.

Campbell, Nick. 2000. Database of emotional speech. *Proceedings ITRW on Speech and Emotion*, Newcastle, 34–38.

Douglas-Cowie, Ellen, Nick Campbell, Roddy Cowie and Peter Roach. 2003. Emotional speech: Towards a new generation of databases. *Speech Communication* 40. 33–60.

Den, Yasuharu. 2009. Prolongation of clause-initial mono-word phrases in Japanese. In Shu-Chuan Tseng (ed.), *Linguistic patterns in spontaneous speech*, 167–192, Taipei: Institute of Linguistics, Academia Sinica.

Den, Yasuharu and John Fry. 2000. Nihongo callhome kōpasu [Callhome Japanese corpus]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 4(2). 24–30.

Den, Yasuharu, Hanae Koiso, Takehiko Maruyama, Kikuo Maekawa, Katsuya Takanashi, Mika Enomoto and Nao Yoshida. 2010. Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. *Proceedings of the 7th International Conference on Language Resources and Evaluation* (LREC 2010), Malta, 2103–2110.

Enomoto, Mika. 2007. Nihongo ni okeru tān kōsei tan'i no ninchi mekanizumu [The cognitive mechanism of the completion of turn-constructional units in Japanese conversation]. *Shakai Gengo Kagaku* [*The Japanese Journal of Language in Society*] 9(2). 17–29.

Fon, Janice. 2002. A Cross-linguistic study on discourse and syntactic boundary cues in spontaneous speech: Using duration as an example. *Proceedings of the International Symposium on Chinese Spoken Language Processing* (ISCSLP 2002), Taipei, 379–382.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Hirai, Toshio, Naoto Iwahashi, Norio Higuchi and Yoshinori Sagisaka. 1995. Tōkeiteki shuhō o mochiita kihon shūhasū seigyo kisoku no jidō chūshutsu [Automatic extraction of fundamental frequency control rules using statistical analysis]. *The Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, 11. 1572–1580.

Hirata, Yukari. this volume. Chapter 18: L2 phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Horiuchi, Yasuo, Yukiko Nakano, Hanae Koiso, Masato Ishizaki, Hiroyuki Suzuki, Michio Okada, Makiko Naka, Syun Tutiya and Akira Ichikawa. 1999. Nihongo chizu kadai kōpasu no sekkei to tokuchō [The design and statistical characterization of the Japanese map task dialogue corpus]. *Jinkō Chinō Gakkaishi* 14(2). 261–272.

Igarashi, Yosuke, Hideaki Kikuchi and Kikuo Maekawa. 2006. Inritsu jōhō [Prosodic information]. In Kokuritsu Kokugo Kenkyūjo (ed.), *Kokuritsu Kokugo Kenkyūjo hōkoku 124: Nihongo hanashi kotoba kōpasu no kōchikuhō* [Research report 124: Construction of the Corpus of Spontaneous Japanese], 347–453. Tokyo: National Institute for Japanese Language.

Igarashi, Yosuke, Ken'ya Nishikawa, Kuniyoshi Tanaka and Reiko Mazuka. 2013. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134(2). 1283–1294.

Ishi, Carlos Toshinori, Hiroshi Ishiguro and Norihiro Hagita. 2010. Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech. *EURASIP Journal of Audio, Speech, and Music Processing*. (doi: 10.1155/2010/528193).

Ishihara, Shunichi. 2010. Variability and consistency in the idiosyncratic selection of fillers in Japanese monologues. *Proceedings of Australasian Language Technology Association Workshop*, 9–17.

Ishizuka, Kentaro, Ryoko Mugitani, Hiroko Kato and Shigeaki Amano. 2007. Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants. *Journal of the Acoustical Society of America* 121(4). 2272–2282.

Itahashi, Shuichi and Chiu-yu Tseng. 2010. *Computer Processing of Asian Spoken Languages*. Los Angeles: Consideration Books.

Itou, Katsunobu, Tomoyasu Akiba, Osamu Hasegawa, Satoru Hayamizu and Kazuyo Tanaka. 1999. A Japanese spontaneous speech corpus collected using automatically inferencing Wizard of OZ system. *Journal of the Acoustical Society of Japan (E)* 20(3). 207–213.

Iwahashi, Naoto and Yoshinori Sagisaka. 2000. Statistical modeling of speech segment duration by constrained tree regression. IEICE Transaction, E83-D, 7. 1550–1559.

Kagomiya, Takayuki, Kenji Yamazumi, Yohichi Maki and Kikuo Maekawa. 2007. Chōshu jikken ni motozuku kōen onsei no inshō hyōtei dēta no kōchiku to sono bunseki [Development and analysis of a psychological evaluating database of public speaking]. *Shakaigengo Kagaku* 9(2). 65–76.

Kagomiya, Takayuki, Kenji Yamazumi, Yohichi Maki and Kikuo Maekawa. 2008. Jihatsu onsei ni okeru taikyokuteki na hatsuwa sokudo no chikaku ni eikyō o ataeru yōin [Factors that affect global perceived speaking rate of spontaneous speech]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 12(1). 54–62.

Kaiki, Nobuyoshi, Kazuya Takeda and Yoshinori Sagisaka. 1991. Gengo jōhō o riyō shita boin keizoku jikanchō no seigyo [Vowel duration control using linguistic information]. *The Transactions of the Institute of Electronics, Information and Communication Engineers A*. 467–473.

Kawakami, Shin. 1963. Bunmatsu nado no jōshōchō ni tsuite [On sentence-final pitch rises]. *Kokugo Kenkyū* 16. 25–46.

Kawatsu, Hiromi and Kikuo Maekawa. 2009. Shiin no chōon yōshiki ga boin no museika ni oyobosu eikyō: Nihongo hanashi kotoba kōpasu no bunseki [Influence of manner of articulation on vowel devoicing: Analysis using the CSJ]. *Nihon Onkyō Gakkai 2009 nen Shunki Kenkyū Happyōkai Kōen Ronbunshū*, 443–444.

Kikuchi, Hideaki and Kikuo Maekawa 2007. Construction of XML documents for the study of prosody using the Corpus of Spontaneous Japanese. *Proceedings Oriental COCOSDA 2007*, Hanoi, 38–42.

Kikuchi Hideaki, Takahiro Miyajima and Raymond Shen. 2013. Tayō na onsei hyōgen kōpasu ni okeru kumatsu onchō no kurasutaringu [Clustering of boundary tones and accentual phrase edge in the expressive speech corpus]. *Proceedings of the 3rd Corpus Japanese Linguistics Workshop*, Tokyo, 23–28.

Kinoshita, Yuko, Shunichi Ishihara and Philip Rose. 2009. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law* 16(1). 91–111.

Kobayashi, Tetsunori, Shuichi Itahashi, Satoru Hayamizu and Toshiyuki Takezawa. 1992. Nihon onkyō gakkai kenkyū yō renzoku onsei dētabēsu [ASJ continuous speech corpus for research]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 48(12). 888–893.

Koiso, Hanae and Yasuharu Den. 2010. A phonetic investigation of turn-taking cues at multiple unit-levels in Japanese conversation. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS XVII), 1122–1125.

Koiso, Hanae and Yasuharu Den. 2013. Acoustic and linguistic features related to speech planning appearing at weak clause boundaries in Japanese dialog. *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* (DiSS 2013), Stockholm, 37–40.

Koiso, Hanae, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa and Yasuharu Den. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech* 41(3–4). 295–321.

Koiso, Hanae, Atsushi Shimojima and Yasuhiro Katagiri. 1998. Collaborative signaling of informational structures by dynamic speech rate. *Language and Speech* 41(3–4). 323–350.

Kori Shiro. 2008. Tōkyō hōgen ni okeru heijobun matsu no kakō zōdai genshō: Heijobun matsu wa heichō ka kakōchō ka [Sentence-final increase of pitch descent in Tokyo Japanese]. *Onsei Gengo* VI. 79–102.

Maeda, Kazuaki and Jennifer J. Venditti. 1998. Phonetic investigation of boundary pitch movements in Japanese. *Proceedings of the 5th International Conference on Spoken Language Processing* (ICSLP 98). Paper #800 of the CD-ROM.

Maekawa, Kikuo. 2003. Corpus of spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition* (SSPR 2003), 7–12.

Maekawa, Kikuo. 2004a. Production and perception of 'paralinguistic' information. *Proceedings of Speech Prosody 2004*, Nara, 367–374.

Maekawa, Kikuo. 2004b. Nihongo hanashi kotoba kōpasu no gaiyō [Outline of the CSJ]. *Nihongo Kagaku* 15. 111–133.

Maekawa, Kikuo. 2009. Analysis of language variation using a large-scale corpus of spontaneous speech. In Shu-Chuan Tseng (ed.), *Linguistic patterns in spontaneous speech*, 27–50. Taipei: Institute of Linguistics, Academia Sinica.

Maekawa, Kikuo. 2010a. Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese. *Journal of Phonetics* 38(3). 360–374.

Maekawa, Kikuo. 2010b. Nihongo yūsei haretsuon ni okeru heisa chōon no jakka [Weakening of stop articulation in Japanese voiced plosives]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 14(2). 1–15.

Maekawa, Kikuo. 2011a. Discrimination of speech registers by prosody. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 1302–1305.

Maekawa, Kikuo. 2011b. PNLP no onseiteki keijō to gengoteki kinō [Phonetic shape and linguistic function of penultimate non-lexical prominence]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 15(1). 1–13.

Maekawa, Kikuo. 2012. Prediction of non-linguistic information of spontaneous speech from the prosodic annotation: Evaluation of the X-JToBI system. *Proceedings of the 8th International Conference on Language Resources and Evaluation* (LREC 2012), 991–996.

Maekawa, Kikuo. 2013a. Kōpasu no sonzai igi [Reasons of existence of corpus]. In Kikuo Maekawa (ed.), *Kōpasu Nyūmon* [Introduction to corpus]. Tokyo: Asakura Shoten.

Maekawa, Kikuo. 2013b. Prediction of F0 height of filled pauses in spontaneous Japanese: A preliminary study. *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* (DiSS 2013), Stockholm, 41–44.

Maekawa, Kikuo. 2013c. Nihongo jihatsu onsei ni okeru final lowering no seiki ryōiki [Domain of final lowering in spontaneous Japanese]. *Proceedings of the 27th Annual Meeting of the Phonetic Society of Japan*, 47–52.

Maekawa, Kikuo and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In Jeroen van de Weijer, Kensuke Nanjo and Tetsuo Nishihara, (eds.), *Voicing in Japanese*, 205–228. Berlin and New York: Mouton de Gruyter.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti. 2002. X-JToBI: An extended J_ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing* (ICSLP 2002), Denver, 1545–1548.

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara. 2000. Spontaneous speech corpus of Japanese. *Proceedings of the Second International Conference on Language Resources and Evaluation* (LREC 2000), 947–952.

Martin, Andrew, Sharon Peperkamp and Emmanuel Dupoux. 2012. Learning phonemes with a proto-lexicon. *Cognitive Science* (DOI: 10.1111/j.1551-6709.2012.01267.x). 1–22.

Maruyama, Takehiko. 2013. *Hatsuwa no jitsujikanteki sanshutsu kara mita hiryūchōsei no kijutsuteki kenkyū* [Descriptive research on disfluent phenomena in spontaneous Japanese: A corpus-based approach]. Tokyo: International Christian University dissertation.

Mazuka, Reiko, Yosuke Igarashi and Ken'ya Nishikawa. 2006. Input for learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus. *Technical report of IEICE*, TL2006-16 106(165). 11–15.

Miyajima, Takahiro, Hideaki Kikuchi and Katsuhiko Shirai. 2011. Collection and analysis of emotional speech focused on the psychological and acoustical diversity. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS 2011), Hong Kong, 1394–1397.

Miyata Susanne (ed.). 2004. *Kyōkara tsukaeru hatsuwa dētabēsu: Shoshinsha no tame no CHILDES nyūmon* [CHILDES for beginners]. Tokyo: Hituzi Syobo.

Mori, Hiroki, Tomoyuki Satake, Makoto Nakamura and Hideki Kasuya. 2011. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics. *Speech Communication* 53. 36–50.

Nakagawa, Natsuko, Yoshihiko Asao and Naonori Nagaya. 2008. Information structure and intonation of right-dislocation sentences in Japanese. *Kyoto University Linguistic Research* 27. 1–22.

Nakagawa, Natsuko, Daisuke Yokomori and Yoshihiko Asao. 2010. The short intonation unit as a vehicle of important topics. *Papers in Linguistic Science* 16. 111–131.

Nakamura, Masanobu, Kōji Iwano and Sadaoki Furui. 2008. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language* 22(2). 171–184.

Nishimura, Ryōta, Norihide Kitaoka and Seiichi Nakagawa. 2009. Onsei taiwa ni okeru inritsu henka o motarasu yōin bunseki [Analysis of factors to make prosodic change in spoken dialog]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 13(3). 66–84.

NLRI (ed.). 1955. *Kokuritsu Kokugo Kenkyūjo hōkoku 8: Danwago no jittai* [NLRI research report 8: Research in the colloquial Japanese]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

NLRI (ed.). 1960. *Kokuritsu Kokugo Kenkyūjo hōkoku 18: Hanashi kotoba no bunkei 1: Taiwa shiryō ni yoru kenkyū* [Research report 18: Sentence patterns in Japanese 1: Research based on conversational speech]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

NLRI (ed.). 1963. *Kokuritsu Kokugo Kenkyūjo hōkoku 23: Hanashi kotoba no bunkei 2: Dokuwa shiryō ni yoru kenkyū* [Research report 23: Sentence patterns in Japanese 2: Research based on monology]. Tokyo: Kokuritsu Kokugo Kenkyūjo.

Oh, Yoon Mi, François Pellegrino, Christophe Coupé and Egidio Marsico. 2013. Cross-language comparison of functional load for vowels, consonants, and tones. *Proceedings of Interspeech 2013*, Lyon, 3032–3036.

Ohsuga, Tomoko, Yasuo Horiuchi, Masafumi Nishida and Akira Ichikawa. 2006. Onsei taiwa de no washa kōtai/keizoku no yosoku ni okeru inritsu jōhō no yūkōsei [Prediction of turn-taking from prosody in spontaneous dialogue]. *Jinkō Chinō Gakkaishi* [*Transactions of the Japanese Society for Artificial Intelligence*] AI 21. 1–8.

Ohta, Kengo, Masatoshi Tsuchiya and Seiichi Nakagawa. 2007. Construction of Spoken Language Model Including Fillers Using Filler Prediction Model. *Proceedings of Interspeech 2007*, Antwerp, 1489–1492.

Ohta, Kengo, Masatoshi Tsuchiya and Seiichi Nakagawa. 2010. Pōzu o kōryo shita hanashi kotoba gengo moderu no kōchiku [Construction of spoken language model considering pause]. *Jōhō Shori Gakkai Ronbunshi* [*Transactions of Information Processing Society of Japan*] 53(2). 889–900.

Oishi, Hatsutaro. 1959. Purominensu ni tsuite: Tōkyōgo no kansatsu ni motozuku oboegaki [On prominence: Notes based upon the observation of Tokyo Japanese]. In Kokuritsu Kokugo Kenkyujo (ed.), *Kotoba no kenkyū* 1, 87–102. Tokyo: Shūei Shuppan.

Ota, Mitsuhiko. this volume. Chapter 17: L1 phonology: phonological development. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.

Sagisaka, Yoshinori. 1993. Inritsu seigyo kenkyū no genjō to kadai: Yori shizen na onsei o motomete [Perspectives on prosody control research in speech synthesis]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 49(12). 854–859.

Sagisaka, Yoshinori and Yoh'ichi Tohkura. 1984. Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyo kisoku [Phoneme duration control for speech synthesis by rule]. *Denshi Tsūshin Gakkai Ronbunshi* [The Transactions of the Institute of Electronics, Information and Communication Engineers A] 67(7). 629–636.

Sagisaka, Yoshinori and Noriyoshi Uratani. 1992. ATR onsei gengo dētabēsu [ATR spoken language database]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 48(12). 878–882.

Sato, Hirokazu. 2011. Eiga ni okeru kaiwa no jikūkan bunseki: Ozu eiga o daizai to shite [Tempo-spatial analysis of conversation in movies: The case of Ozu's movies]. In Miyoko Sugito (ed.), *Onsei bunpō* [Speech and grammar], 151–168. Tokyo: Kurosio.

Takeda, Kazuya, Yoshinori Sagisaka and Hisao Kuwabara. 1989. On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America* 86(6). 2081–2087.

Venditti, Jennifer. 1997. Japanese ToBI Labelling Guidelines. In Kim Ainsworth-Darnell and Mariapaola D'Imperio (eds.), *Papers from the Linguistics Laboratory.* Ohio State University Working Papers in Linguistics 50. 127–162. (First distributed in 1995 as a web document).

Venditti, Jennifer J., Kikuo Maekawa and Mary M. Beckman. 2008. Prominence marking in the Japanese intonation system. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 456–512. Oxford: Oxford University Press.

Wada, Toshiya, Takahiro Shinozaki and Sadaoki Furui. 2010. Investigations of features and estimators for speech-based age estimation. *Proceedings of the Second APSIPA Annual Summit and Conference*, Singapore, 470–473.

Warner, Natasha and Takayuki Arai. 2001. The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109(3). 1144–1156.

Watanabe, Michiko. 2010. *Features and roles of filled pauses in speech communication*. Tokyo: Hituzi Syobo.
Watanabe, Michiko. 2013. Phrasal complexity and the occurrence of filled pauses in presentation speeches in Japanese. *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* (DiSS 2013), Stockholm, 71–71.
Watanabe, Michiko, Keikichi Hirose, Yasuharu Den and Nobuaki Minematsu. 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication* 50(2). 81–94.
Yamazumi, Kenji, Takayuki Kagomiya, Yohichi Maki and Kikuo Maekawa. 2005. Kōen onsei no inshō hyōka shakudo [Psychological scale for the impression rating of monologue]. *Nihon Onkyō Gakkaishi* [Journal of the Acoustical Society of Japan] 61(6). 303–311.
Yoneyama, Kiyoko, Hanae Koiso and Janice Fon. 2003. A Corpus-based analysis on prosody and discourse structure in Japanese spontaneous monologues. *Proceedings of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition* (SSPR), Tokyo, 27–30.

Mitsuhiko Ota

17 L1 phonology: phonological development

1 Introduction

This chapter presents an overview of descriptive findings in the phonological acquisition of Japanese as a native language, and discusses their implications for our understanding of the phonological structure of Japanese, general phonological theory and models of phonological development. The selection of topics largely reflects the availability of developmental research on each area, but an effort has been made to give due attention to areas that are likely to be of great interest to theoretical phonologists even when the acquisition literature is rather sparse. Furthermore, certain topics are addressed in detail in order to explore their relevance to general issues in phonological development. The chapter also highlights specific findings from the acquisition of Japanese that complement previous research in phonological acquisition, which has been built primarily on data from major European languages, especially English.

The chapter is organized as follows. Section 2 is devoted to topics on segmental development, including early segmental perception and production, order of acquisition, substitutions, vowel devoicing and phonotactics. Section 3 reviews the developmental literature on duration-based contrasts, i.e., short vs. long vowels and singleton vs. geminate consonants. Section 4 describes the acquisition of pitch phonology, that is, pitch accent, intonation and their interaction. Section 5 examines children's word production and speech segmentation data in relation to the question of when and how words in Japanese begin to show internal organization in terms of prosodic units such as moras, syllables and feet. Section 6 discusses the development of *rendaku* voicing and lexical stratification, two topics that are related to the acquisition of phonology in connection with morphology and the lexicon. The chapter concludes with suggestions for future directions.

2 Segmental development

2.1 Early perception and production

In stark contrast to the great number of studies carried out on the development of sound production in Japanese-learning children over the age of one year, there is a noticeable paucity of developmental research on how younger infants perceive or produce phonetic differences related to segmental contrasts. One exception to this is the work related to the perceptual decline in the discrimination of the English

sounds [r] and [l]. At 6 to 8 months of age, Japanese-learning infants can discriminate English [r] and [l] at a level comparable to that of infants exposed to American English, but by 10 to 12 months, their performance becomes significantly lower than their American counterparts' (Kuhl et al. 2006). This observation is consistent with results from other languages, in which a similar loss of sensitivity to nonnative contrasts has been observed by 10 months, indicating that infants' auditory perception has become more attuned to specific acoustic differences in the phonemic contrasts of the ambient language (Best and McRoberts 2003; Best et al. 1995; Werker and Tees 1984). As with infants exposed to other languages, some time between 8 and 10 months of age, Japanese-learning infants begin to "tune out" some of the phonetic differences that are not phonemically relevant to their language.

Evidence for attunement to the language-specific distribution of segments in Japanese can also be found in prelinguistic production. Boysson-Bardies and Vihman (1991) compared babbling produced by children learning English, French, Japanese and Swedish. The Japanese infants (from 13 to 19 months) produced significantly fewer labials (25.4%) than English or French infants, more stops (69.5%) than French infants, and more nasals (16.2%) than Swedish infants. These differences in the composition of babbling sounds mirror the distribution of place and manner categories in the adult languages: Japanese has a lower proportion of labials than English or French, a higher proportion of stops than French, and a higher proportion of nasals than Swedish. Thus, even in babbling, the phonetic production of Japanese-exposed infants shows an influence of the frequency distribution of segmental sounds from the linguistic input.

2.2 Order of segmental development in production

Once children begin producing words with identifiable adult correspondents, it becomes possible to track how the production of individual segments approximate their adult targets over time. This was the aim of the many norming studies carried out primarily in the 1960s and 1970s (Murata 1970; Nakajima et al. 1962; Nakanishi 1982; Nakanishi, Ōwada, and Fujita 1972; Noda et al. 1969; Ōwada and Nakanishi 1971; Ōwada, Nakanishi, and Oshige 1969; Sakauchi 1967; Takagi and Yasuda 1967; Umebayashi and Takagi 1965). Most of these studies used structured elicitation (e.g., picture description) to obtain cross-sectional average scores in production accuracy (see Ota and Ueda 2006 for an overview). Despite the considerable amount of variability across studies and individuals, several general patterns emerge from this

1 For Noda et al. (1969), Nakanishi, Ōwada, and Fujita (1972) and Sakauchi (1967), the age range indicates the youngest cross-sectional group that met the 90% correct criterion. For Nakanishi (1982), it shows (along with the mean given in brackets) when the sound was first produced and completely mastered in longitudinal data.

Table 1: Age of acquisition for word-initial consonants

	Noda et al. (1969)	Nakanishi et al. (1972)	Sakauchi (1967)	Nakanishi (1982)	
				First appearance	Complete acquisition
m	3;0-3;5	4;0-4;5	2;10-3;3	1;0-1;8 (1;3)	1;10-2;9 (2;2)
b	3;0-3;5	4;0-4;5	2;10-3;3	1;0-1;6 (1;3)	1;3-2;9 (2;3)
p	3;6-3;11	4;0-4;5	2;10-3;3	1;2-1;9 (1;5)	1;8-2;6 (2;1)
t	3;0-3;5	4;0-4;5	2;10-3;3	1;1-1;10 (1;5)	1;10-2;9 (2;3)
j	3;0-3;5	4;0-4;5	2;10-3;3	1;1-2;0 (1;5)	1;4-4;0 (2;5)
ʧ	3;0-3;5	4;0-4;5	2;10-3;3	1;2-1;10 (1;6)	1;8-2;9 (2;5)
k	3;6-3;11	4;0-4;5	2;10-3;3	1;1-1;10 (1;5)	1;10-2;6 (2;2)
g	3;6-3;11	4;0-4;5	2;10-3;3	1;3-2;3 (1;9)	1;7-2;9 (2;3)
n	4;0-4;5	4;0-4;5	3;4-3;8	1;1-1;9 (1;4)	1;10-2;3 (2;0)
d	4;6-4;11	4;0-4;5	2;10-3;3	1;2-1;9 (1;5)	1;11-3;0 (2;5)
ʤ	3;6-3;11	4;0-4;5	3;4-3;8	1;2-2;3 (1;8)	1;11-4;0 (2;8+)
w	4;6-4;11	4;0-4;5	3;4-3;8	1;1-1;9 (1;5)	1;6-4;0 (2;7)
h	4;0-4;5	4;0-4;5	2;10-3;3	1;2-1;8 (1;6)	2;6-4;0 (3;1)
ɸ	4;0-4;5	4;0-4;5	2;10-3;3	n/a	n/a
ʃ	4;0-4;5	4;0-4;5	after 4;8	n/a	n/a
ɾ	4;0-4;5	5;6-5;11	after 4;8	1;3-2;3 (1;8)	2;0-4;0 (3;3+)
ç	5;6-5;11	4;6-4;11	4;4-4;8	1;1-2;3 (1;8)	2;9-4;0 (3;4+)
s	5;0-5;5	5;0-5;5	after 4;8	1;5-2;9 (2;3)	3;0-4;0 (4;0+)
z	5;6-5;11	5;6-5;11	after 4;8	1;7-3;0 (2;6)	3;0-4;0 (3;6+)
ʦ	5;6-5;11	5;0-5;5	after 4;8	1;6-3;0 (2;8)	3;0-4;0 (3;8+)

body of research. Performance thresholds are met earlier in vowels than in consonants, with the majority of children reaching the 90% criterion for all 5 basic vowels by the age of 2 years. Among the consonants, stops and nasals are acquired early (mostly by age 4), while fricatives, particularly sibilants ([s, z, ʃ]), and the flap ([ɾ]) are acquired later. Some representative data for word-initial consonants are summarized in Table 1.

The order of development presented here parallels the overall pattern attested in other languages. Crosslinguistically, target-like production tends to be attained earlier in stops and nasals than in fricatives, affricates and liquids; among fricatives, sibilants are generally acquired the latest (Ingram 1989; Kent 1992; Smit et al. 1990; Chevrie-Muller and Lebreton, 1973). In feature-building models of segmental development, this general order is seen as a manifestation of markedness in feature values (Brown and Matthews 1997; Rice and Avery 1995). Under this view, fricatives are not acquired before stops because [+continuant] is more marked than [-continuant], and sibilants are not acquired before non-sibilants because [+strident] is more marked than [-strident] (Ueda 1996). Such markedness generalizations are, of course, likely

to be phonetically grounded in the relative articulatory demands for different sounds. For example, the articulatory gestures involved in creating the narrow constriction for fricatives are motorically more demanding than the simple ballistic movement that produces stops, which can explain why the mastery of fricatives is slower than that of stops (Kent 1992).

Table 1 shows that this putatively universal order of development is disrupted in Japanese by the sibilant affricates [ʧ] and [ʤ]. Contrary to the crosslinguistically prevalent pattern in which sibilant affricates are acquired much later than corresponding stops, [ʧ] and [ʤ] in Japanese typically reach production accuracy criteria around the same time as their stop counterparts [t] and [d]. However, a closer analysis by Edwards and Beckman (2008a) reveals that this non-compliance of the universal order is conditioned by the subsequent vowel. In most contexts, [ʧ] actually lags behind [t] in development. But [ʧ] has a higher production accuracy than [t] before /i/, making the average timing of acquisition similar for the two sounds. The pre-/i/ context is where the contrast between /t/ and /ʧ/ neutralizes to [ʧ] in Japanese except in some loanwords (e.g., [paːtiː] 'party'), and consequently presents substantially more instances of [ʧ] than [t] in the input. The relatively early acquisition of sibilant affricates in Japanese, therefore, can be seen as a case where a language-specific frequency effect is "overlaid on a universal articulatory ease effect" (Edwards and Beckman 2008a: 146).

A caveat in all the production studies cited above is that the results are based on phonetic transcription or researchers' judgments of the adultlikeness of the production. This type of data is a valid indicator of how the child's production is perceived by mature members of the speech community, but may miss reliable acoustic differences that are imperceptible to adult speakers (Edwards and Beckman 2008b, Baum and McNutt 1990; Macken and Barton 1980; Maxwell and Weismer 1982). Indeed, when Li, Edwards, and Beckman (2009) carried out an acoustic analysis of voiceless sibilants produced by 2- and 3-year-old Japanese children, they uncovered subtle differences between the sounds targeting [s] and [ʃ], even though the transcription analysis suggested no differentiation (note in Table 1, [s] and [ʃ] are listed as having an age of acquisition typically after 4). In some cases the difference was found in the peakiness of the spectral distribution (reflecting the more compact tongue posture of [ʃ]-targeting sounds), and in others, in the height of the second formant at the onset of the following vowel (reflecting the shorter back cavity created by [ʃ]-targeting sounds). Thus, although contrasts such as [s] – [ʃ] may not sound adult-like until 4 or 5 years of age, they begin to show fine phonetic differentiation by 2 or 3 years.

2.3 Substitutions

Children's early segmental production often exhibits phonetic drifts that result in a sound that is perceived by adult listeners as similar or identical to a different segment in the language. The mechanisms that underlie such substitution patterns are

best understood by comparing the similarities and differences across languages. Commonly found patterns are likely to reflect some general phonetic or phonological factors that constrain the development of segmental production, while cross-linguistically divergent patterns should reveal the effects of subtle articulatory differences in the equivalent target sounds or the relationship between the target sound and other segments in the language.

One type of substitution errors typically found in Japanese, as well other languages, relates to fricatives. As illustrated in (1), Japanese-speaking children often produce sibilant fricatives with complete or near-complete closure (Masuko 2004; Ōwada, Nakanishi, and Ōshige 1969; Ōkubo 1977; K. Itō 1990).

(1) Stopping ([s, ʃ] → [t, ʧ])[2]
   a. [sakana] → [takana] ‘fish’
   b. [kiʃa] → [kita] ‘train’
   c. [hoʃiː] → [hoʧiː] ‘I want it’

This is a pattern consistent with the so-called ‘stopping’ of fricatives that is widely attested in other languages (e.g., *see* → [ti], *zoo* → [du]; Smith 1973; Locke 1983). The most likely explanation for this substitution pattern is the relative articulatory difficulty involved in the production of fricatives (Vihman 1996). To produce a fricative, the child needs to position the articulator so that a narrow constriction is formed to create turbulence in the airflow. When motor control skills are still in development, attempts at this articulatory gesture can be imprecise or overcommitted, resulting in a complete closure, or a stop.

However, substitutions of sibilants in children acquiring Japanese do not always follow the same patterns as those in children acquiring other languages. One typical pattern of sibilant substitution found in Japanese-learning children is the production of target alveolar sibilants ([s], [z], [ts]) with a more palatal articulation, yielding sounds that are perceived as [ʃ, ʧ, ʒ]) (Ōwada, Nakanishi, and Ōshige 1969; Ōkubo 1977; K. Itō 1990). Some examples of such “palatalization” are given in (2). The overall pattern is the opposite in English-learning children, who typically produce a target palatal fricative /ʃ/ as a more anterior sound that is perceived as [s] (*shoe* → [su], *ship* → [sɪp], (Li, Edwards, and Beckman 2009; Weismer and Elbert 1982).

(2) Palatalization (/s, ʦ, z/ → [ʃ, ʧ, ʒ])
   a. [ɯsagi] → [ɯʃagi] ‘rabbit’
   b. [mizɯ] → [miʒɯ] ‘water’
   c. [ʦɯmiki] → [ʧɯmiki] ‘block’

2 Correspondence between adult target forms and children’s productions is shown in the format “[adult phonetic form] → [child’s phonetic form]”.

A potential source of this crosslinguistic difference is the articulatory details of the sibilant sounds in these languages. In English, the difference between [s] and [ʃ] is primarily a matter of tongue position with the latter having a more posterior constriction (Fletcher and Newman 1991; Stone et al. 1992). In contrast, [s] and [ʃ] in Japanese are differentiated by tongue posture rather than position, with the latter having a longer constriction (Akamatsu 1997; Li, Edwards, and Beckman 2009). This subtle difference is reflected in the fact that the voiceless palatal fricative in Japanese is often transcribed as [ɕ] (alveolo-palatal) instead of [ʃ] (palato-alveolar) (Okada 1999; Vance 2008). The motor requirements in producing a sibilant may be higher when the constriction is further back in the oral cavity than the alveolar region and, likewise, when it is narrower in tongue contact, thus placing articulatory pressure on English [ʃ] toward [s] and Japanese [s] toward [ʃ] (or [ɕ]).

Another substitution pattern in Japanese that warrants crosslinguistic comparisons is related to the intended production of the liquid sound ([ɾ]). Though occasionally produced as a palatal glide ([j]), [ɾ] is more commonly substituted with a sound with a complete oral closure, or [d] (Umebayashi and Takagi 1965; Sakauchi 1967; Murata 1970; Ueda, Ito, and Shirahata 1998). In some cases, all instances of target [ɾ] are realized as [d] regardless of their position. In other cases, [d] and [ɾ] stand in a complementary distribution pattern, whereby [d] tends to occur word-initially and [ɾ] word-medially, as illustrated in (3).

(3) A specific substitution pattern for target [ɾ] and [d] (Ueda, Itō and Shirahata 1997)
a. [ɾappa] → [dappa] 'trumpet'
b. [denʃa] → [denʃa] 'train'
c. [teɾebi] → [teɾebi] 'TV'
d. [bɯdoː] → [bɯɾoː] 'grape'

Substitution errors of liquids typically result in different sounds in other languages. For example, English [ɹ] and [l] in word-initial position are most commonly produced as a glide (e.g., *red* → [wɛd], Smit 1993). The three liquid sounds in Spanish, [ɾ], [l] and [r] (trill) are often substituted with each other (Goldstein, Fabiano, and Washington 2005). These observations indicate that substitution of liquids is dependent both on the phonetic characteristics of the target sound as well as its proximity to other sounds in the inventory of the language. The Japanese liquid is a flap, which is produced by a strike of the tongue against the alveolar ridge (Ladefoged 1971). To the extent that this yields an oral closure (albeit an extremely brief one), an alveolar flap is similar to an alveolar stop ([d]). This accounts for the general substitution of [d] for [ɾ] and also the complementary distribution pattern illustrated in (3) since the utterance-initial position, where air pressure can build up prior to articulation, is more conducive to a release burst that is characteristic of [d]. In fact, there is some indication that adult pronunciation of [ɾ] in utterance-initial position also involves a

weak but complete closure (Kawakami 1977; Vance 1987). English liquids, on the other hand, are essentially approximants that share more articulatory similarities with glides than with stops.

There are other typical substitution patterns found in child Japanese that also differ slightly from crosslinguistic patterns, but without any immediately obvious articulatory reasons. For instance, the substitution of the velar stop [k] by the alveolar [t] is frequently documented in children learning English, and has even been suggested to be a manifestation of universal markedness against back consonants, which induces "fronting" errors (Ingram 1974; Locke 1983). Contrary to this claim, Japanese-speaking children produce both "fronting" errors, as given in (4), and the opposite "backing" errors, as given in (5) (Nakanishi, Ōwada, and Fujita 1972; Beckman, Yoneyama, and Edwards 2003).

(4) Fronting errors (Ueda 1996)
  a. [mikaɴ] → [mitaŋ] 'tangerine'
  b. [poketto] → [potetto] 'pocket'
  c. [kaɲi] → [taɲi] 'key'

(5) Backing errors (Beckman, Yoneyama, and Edwards 2003)
  a. [toɾa] → [koɾa] 'tiger'
  b. [tanɯki] → [kikːi] 'badger'

Although Japanese alveolar stops are said to be more laminal than those in English, French or Spanish (Vance 2008), it is not entirely clear how this causes the crosslinguistic difference in substitutions. It has been suggested that the contrast in fronting and backing errors is due to the frequency distribution of alveolar and velar stops in the ambient language (Beckman, Yoneyama, and Edwards 2003; Mazuka 2010). In many languages, including English, /t/ occurs more frequently than /k/, providing the learner with more exposure and opportunities to produce /t/. This conspires with the claimed markedness effect that favors /t/ over /k/ (Bernhardt and Stemberger 1998). Conversely, in Japanese, /k/ occurs more frequently than /t/ both in adult corpora and lexical items that are more likely to be heard in child-directed speech (Beckman, Yoneyama, and Edwards 2003). It is therefore possible that there is a universal phonetic effect that induces fronting, but the effect is counteracted in the case of Japanese acquisition by the input distribution that is skewed more toward /k/ rather than /t/.

2.4 Segmental processes and phonotactics

A well-known segmental process in Tokyo Japanese is the devoicing of short high vowels (/i/ and /u/) between two voiceless obstruents or between a voiceless obstruent

and a pause (see Fujimoto, this volume, for an overview). Very little is known about the development of vowel devoicing in young children except that it affects their word production. Children of 1 or 2 years of age frequently omit syllables that contain a devoiced vowel as shown in (6) (Ota 2003a).

(6) Omission of syllables with devoiced vowels

a.	[naiɸɯ̥] → [nai] ‘knife’	(Hiromi 1;10.11)[3]
b.	[kɯ̥tsɯʃi̥ta] → [tsutta] ‘socks’	(Hiromi 1;11.9)
c.	[ʒɯːsɯ̥] → [ʤuː] ‘juice’	(Takeru 1;8.13)
d.	[kɯ̥ʧi] → [ʧi] ‘mouth’	(Takeru 1;11.2)
e.	[ki̥ʃa] → [da] ‘train’	(Kenta 2;4.15)

It is possible that this reflects a perceptual effect due to the low amplitude of devoiced vowels, which are often deleted in casual speech especially after fricatives and affricates (Vance 1987). Adults are known to employ coarticulation and timing cues to retrieve the missing vowels, but young children may not have acquired such compensatory strategies.

Adult-like production of vowel devoicing is acquired relatively late, around the age of 5. In an elicited production task, Imaizumi, Fuwa, and Hosoi (1999) measured the vowels produced by 4-year-old, 5-year-old and adult speakers of Tokyo Japanese and Osaka Japanese (the latter being a dialect with no vowel devoicing). The 5-year-old and adult speakers of Tokyo Japanese differed from all the other groups in having a lower prominence and longer duration for the vowels in devoicing contexts, indicating that the 5-year-olds, but not the 4-year-olds, exposed to Tokyo Japanese have converged on the adult speakers in this respect.

As mentioned above, devoicing sometimes leads to deletion of vowels, creating a form that is phonotactically impossible if interpreted without regard to the underlying process (e.g., /batsu/ → [bats] ‘punishment’). This raises an interesting question for phonological acquisition. Would infants exposed to Tokyo Japanese be led to accept phonetic forms such as [bats] as phonotactically legal, or would they show some sensitivity to its noncanonical structure? This issue was addressed in two studies that explored Japanese infants’ sensitivity to word-level phonotactics (Kajikawa et al. 2006; Mugitani et al. 2007). The results showed that 6-month-olds were incapable of discriminating a phonotactically possible but phonetically noncanonical form ([kiːts], which could be a devoiced rendition of /kiːtsu/) from a phonotactically impossible form ([kiːt]) or from a phonotactically possible and phonetically canonical form ([kiːtsɯ]). In contrast, 12-month-olds and 18-month-olds can discriminate the possible but noncanonical [kiːts] from the possible and canonical [kiːtsɯ], but still not from the impossible [kiːt]. As a comparison, English-learning

3 Children’s ages are indicated in the format “years; months.days”.

18-month-olds could discriminate [nik] from [niks], both of which are phonotactically legal in English. These results can be interpreted as evidence that the phonetic difference between two forms due to vowel devoicing is too subtle for 6-month-olds to detect, and that sensitivity to phonotactics in devoiced forms (e.g., [kiːts]) does not develop before the age of 18 months.

3 Duration-based phonemic contrasts

Japanese has duration-based phonemic contrasts in both vowels (e.g., /to/ 'door' vs. /too/ 'ten') and consonants (e.g., /saka/ 'slope' vs. /sakka/ 'writer' and /ama/ 'nun' vs. /amma/ 'masseur'). The latter can also be construed as a distinction between singleton onsets and geminates, but for both, the contrast is cued primarily by the duration of the segment (Vance 2008; see also Kawahara, Ch. 1, this volume, and Kawagoe, this volume, for the phonetics and phonology of geminate obstruents as well as Hirata, this volume, for contrasts in vowel and consonant length in second language acquisition). The following subsections begin with a review of what is known about the developmental timing of these contrasts, first, in perception, then, in production. This is followed by a discussion of the learning mechanisms that may be responsible for the acquisition of these contrasts.

3.1 Perception of durational contrasts

The perceptual development of vowel durations in Japanese-learning infants has been examined through experiments using the visual habituation-dishabituation paradigm (Mugitani et al. 2009; Sato, Sogabe, and Mazuka 2010a). In these experiments, infants are repeatedly exposed to novel word forms that contain either a typical short vowel or long vowel in the critical position (e.g., [mana] vs. [maːna]) until their visual fixation to an accompanying screen image (a checker board pattern) decreases to a predetermined threshold. A recovery in looking time when the auditory word switches to one differing only in the length of the critical vowel indicates infants' ability to detect the difference. The results show that 4-month-olds and 7.5-month-olds do not discriminate between the vowels (Sato, Sogabe, and Mazuka 2010a), while 9-month-olds and 10-month-olds do (Mugitani et al. 2009; Sato, Sogabe, and Mazuka 2010a). In Sato, Sogabe, and Mazuka (2010a), these results were obtained using both naturally produced tokens that contained other cues that covary with duration (such as pitch contours), and digitally manipulated tokens that lacked any secondary cues. The indication is that Japanese-learning infants develop the ability to discriminate vowel length distinctions based solely on durational cues between 7.5 and 9 months of age.

Using the same paradigm, Sato, Kato, and Mazuka (2012) tested Japanese-learning infants' ability to discriminate short (singleton) and long (geminate) consonants in novel word forms such as [pata] and [patta]. When naturally produced tokens were used, the results were similar to the vowel length findings in that 9.5- and 11.5-month-olds could detect the difference but 4-month-olds could not. However, only the 11.5-month-olds could discriminate manipulated tokens that lacked secondary cues, such as the duration of the preceding vowel and intensity differences, and neither the 9.5- or 11.5-month-olds could discriminate manipulated tokens that contained contradictory durational and covarying cues. Taken together with the findings from vowel length perception experiments, these findings suggest that Japanese-learning infants' ability to discriminate duration-based contrasts without secondary cues develop slightly later for consonants (after 9 months) than for vowels (before 9 months).

This overall timing of development for durational contrasts has also been corroborated by neural evidence. Using near-infrared spectroscopy, Minagawa-Kawai et al. (2007) tested infants' neural response to stimuli consisting of a 4-step [mama]-[mamaː] continuum, in which the second (184ms) and third (217ms) steps crossed the typical perceptual boundary in adult Japanese speakers. Even the youngest group of infants (3- to 4-month-olds) showed responses to stimulus changes, but it was only from 6–7 months that significantly stronger responses were recorded to an across-category change (i.e., Step 2 to 3) than to a within-category change in stimuli (Step 1 to 2, Step 3 to 4). Furthermore, the leftward lateralization that is characteristic of adult Japanese listeners (but not of adult listeners of languages without durational vowel contrasts) was only exhibited by infants older than 12 months. It appears, then, that durational differences in vowels are first processed by a general auditory circuit, which is then handled by a more linguistically specialized circuit in the second half of the first year of life.

3.2 Production of durational contrasts

Turning now to production data, studies show that the distinction between short and long segments emerges early in children's word production, but the actual phonetic values of the contrasts take several years to converge on those of adults. For instance, Ota (2003a) measured the duration of vowels in words spontaneously produced by children between the ages of 16 to 19 months (e.g., [çikoːki] 'airplane' vs. [koko] 'here') and found a significant difference between short and long vowels. However, the long vowels were only 1.33 to 1.78 times longer than their short counterpart – a ratio that falls short of the average adult value reported in the literature, which ranges from 1:1.7 to 1:2.1 (Hoequist 1983a, 1983b, Warner and Arai 2001). Similarly, Kunnari, Nakai, and Vihman (2001) examined the duration of stops spontaneously produced by children around 17.5 months, and found that long stops

(geminates) were on average only about 1.48 times the duration of short stops (singleton), also not matching the typical ratio of 1:2 to 1:3 reported in adult production (Han 1992). The 3-, 4-, and 5-year-olds studied in Aoyama (2001) all showed a significant durational difference in their elicited production of singleton [n] and geminate [nn] nasals, as well as a gradual increase in the durational ratio of the geminate nasal with respect to the singleton nasal from 1.37 (3-year-olds) to 1.55 (4-year-olds) and then to 1.59 (5-year-olds). Nonetheless the 5-year-olds' ratio did not match the value recorded by the adult controls (2.07).

The development of durational contrast production in Japanese follows the general pattern attested in many other contrasts, such as the voiced/voiceless distinction in English (Macken and Barton 1980; Edwards and Beckman 2008b), where children initially mark the difference between phonologically contrastive sounds by using non-adultlike phonetic values. It also exhibits the commonly observed perception-production gap in contrast development; that is, evidence of perceiving the contrast can be observed before the age of 1 year, along with a fairly adult-like perceptual boundary of that contrast, but an adult-like phonetic accuracy in production develops much later. The perception and production of durational contrasts in Japanese, therefore, adds to the growing body of literature which suggests that producing a contrast in an adult-like manner requires more than having accurate perceptual categories for the contrasting sounds.

3.3 Learning mechanisms

The discussion in the previous subsection has not directly answered the question of how a child exposed to Japanese comes to the understanding that the language contains durational contrasts. At first blush, this seems like a trivial issue. The child encounters a lexical pair such as /e/ 'picture' and /eː/ 'yes', and deduces that a difference in duration is phonemic in the language. There are two difficulties with this explanation. First, the perception of this contrast appears to emerge too early to be dependent on such lexical evidence. Minimal pairs of the sort /e/ and /eː/, or even near minimal pairs, are few and far between in the lexicon of 9-month-olds, who are already beginning to show linguistic understanding of this contrast. Second, the contrastive nature of durational differences in Japanese needs to be identified amidst other types of variability in segmental duration that reflect intrinsic segmental differences (e.g., high vowels are longer than low vowels), prosodic structure (e.g., phrase-initial and utterance-final segments are longer than others), or an orthogonal contrast (e.g., voiceless stops have longer closure times than voiced ones). The complexity of this task can be demonstrated by imagining a child who is exposed to either English or Japanese. In English, a vowel contrast such as /ɛ/-/e/ is signaled primarily by spectral cues. Yet there is also a systematic difference in duration. In Japanese, a similar contrast, /e/-/eː/ is signaled by durational cues, but at

the same time these sounds also differ in other respects, including formant structure (Hirata and Tsukada 2009). How does a Japanese-learning infant know duration is the phonetic correlate that matters?

One recently proposed solution to this learning problem is the idea that the learner is not comparing the phonetic information between minimally differing words, but instead tracking the overall distribution patterns of acoustic information in the input (Maye, Werker, and Gerken 2002; Maye, Weiss, and Aslin 2007). For example, because of the durational contrast, high-front vowels in Japanese may show a very strong bimodal distribution in duration (i.e., their durations cluster closely around two distinct values), while they show less pronounced distributional patterns in other dimensions. If it is computationally possible to use this type of information to select the correct phonetic dimension for the contrast, then it is plausible that infants exposed to Japanese can learn durational contrasts only by tracking the acoustic characteristics and frequencies of the sounds they hear in the environment. These possibilities have been examined by Werker et al. (2007) and Vallabha et al. (2007).

In Werker et al. (2007), Japanese-speaking mothers and English-speaking Canadian mothers were asked to "teach" novel words provided in a picture book to their 12-month-old infants. The novel words contained pairs contrasting in /e/-/eː/ and /i/-/iː/ for Japanese and /ɛ/-/e/ and /ɪ/-/i/ for English. Acoustic analysis showed that each of the two vowel pairs differed more in duration in the speech produced by the Japanese mothers but more in spectral profile in the speech produced by the English-speaking mothers. Furthermore, a hierarchical regression model showed that better predictions of the category membership of the vowels could be obtained from durational cues in the Japanese data but from spectral cues in the English data. Thus, there is a reliable amount of distributional evidence in the phonetic information of vowels to allow potential learners of Japanese to deduce the duration-based categories.

Vallabha et al. (2007) addressed the question of whether the distributional information uncovered in Werker et al. (2007) is sufficient to learn vowel duration contrasts (or vowel quality contrasts) without knowing beforehand how many categories are to be learned or to which category each token belongs. The simulations run by Vallabha et al. (2007), in fact, successfully learned the contrasts under such conditions (although to different degrees depending on the learning assumptions built into the algorithms), the implication being that the Japanese durational contrast is technically learnable based only on distributional information without a priori language-specific expectations for the relevant phonetic cues.

While the approach based on phonetic input distribution leaves many empirical questions to be answered, it provides a promising alternative to the traditional account of contrast learning based on lexical contrast, an account that has been made less likely by the observed mismatch in the timing of phonological and lexical development.

4 Pitch phonology

4.1 Background

The pitch phonology of Japanese contains two components: a lexical component that assigns a fixed tonal height or contour (pitch accent) to a lexical item, and an intonational component that assigns grammatically- or pragmatically-defined contour patterns to phrases and utterances (see Kawahara, Ch. 11, this volume, and Igarashi, this volume, for detailed descriptions; see also Hirata, this volume, for the discussion of pitch accent in L2 phonology). In Tokyo Japanese, words are either accented or unaccented, and accented words have a lexically-specified position marked by a downfall in pitch. Intonational pitch features in Tokyo Japanese include the lowering of (short) syllables at the beginning of a prosodic phrase. Pitch accent and phrase-initial lowering generate the following contour patterns in the citation forms of disyllabic words (see (7)). The acute accent diacritic indicates the position of the lexical pitch accent.

(7) Pitch contours in Tokyo Japanese

a. hási 'chopstick'

b. ano hási-ga 'that chopstick (NOM)'

c. hasí 'bridge'

d. ano hasí -ga 'that bridge (NOM)'

e. hasi 'edge'

f. ano hasi-ga 'that edge (NOM)'

The learning task that the child faces is to unravel the different phonological components from the composites of pitch accent and intonational patterns in the input data. For example, a child must learn that the contour difference between (7a), (7c) and (7e) is lexically relevant, that there is a difference between the accentual properties of words in (7c) and (7e) despite the surface similarities, and that the rising contour in (7e) is not part of the lexical property of the word *hasi* ('edge').

4.2 Perception of pitch phonology

The perceptual precursors to this learning process are already evident immediately after birth. Neonates exposed only to French detect a change between two lists of

disyllabic Japanese words, one with a falling contour (e.g., *áme* 'rain', *ísi* 'intent', *kámi* 'god') and another with a rising contour (e.g., *ame* 'candy', *isí* 'stone', *kamí* 'paper') (Nazzi, Floccia, and Bentoncini 1998). Similarly, 4- to 16-week-old infants in the US (presumably exposed only to English) detect a change between two lists of synthesized vowels (/a/ and /i/), one with a falling contour (112 to 92Hz) and the other with no change in the fundamental frequency (112Hz) (Kuhl and Miller 1982). Findings such as these indicate that young infants are not only sensitive to low-level acoustic differences in pitch (e.g., Karzon 1985; Karzon and Nicholas 1989), but also able to extract the common pitch characteristics in novel auditory stimulus sets.

Given that words with different pitch contours can be discriminated even by infants who are not learning a language that uses pitch for lexical contrasts, it is not surprising that Japanese-learning infants can do exactly the same. Using a subset of the stimuli from Nazzi, Floccia, and Bentoncini (1998) in a visual habituation-dishabituation paradigm, Sato, Sogabe, and Mazuka (2010b) tested 4-month-old and 10-month-old infants exposed to Tokyo Japanese and demonstrated that both age groups could discriminate between disyllables with falling versus rising contours. However, a separate experiment using near-infrared spectroscopy indicated a difference in the way the two groups processed the pitch contours (Sato, Sogabe, and Mazuka 2010b). When the falling and rising contours were presented in pure tone, the change in pitch induced bilateral hemodynamic responses (i.e., similar activation in the left and right hemispheres of the brain) in both 4-month-olds and 10-month-olds. In contrast, when the contours were embedded in the words used in the habituation experiment, the activation was higher in the left hemisphere than in the right hemisphere for 10-month-olds, but not for 4-month-olds. Left lateralization is generally a sign that a phonetic difference is processed as linguistically relevant, and it is the pattern shown in adult Japanese speakers processing lexical pitch contrasts. Similar to the case of durational contrasts (Minagawa-Kawai et al. 2007), the results from Sato, Sogabe, and Mazuka (2010b) suggest that the perception of pitch contours by Japanese-learning infants undergoes reorganization by 10 months during which it shifts from a general auditory to a more language-specific mode of processing.

4.3 Production of pitch phonology

The language-specific ability to process pitch patterns discussed in the preceding section seems to be employed immediately in the acquisition of lexical pitch. In spontaneous speech produced by 1-year-olds learning Japanese, Hallé, Boysson-Bardies, and Vihman (1991) observe that most isolated initially-accented disyllables (see (7a)) were produced with a global falling contour (67–97%). Then, by this stage, children have at least internalized the falling pitch pattern associated with pitch accent as part of the lexical representation of words they have learned. However,

not all aspects of pitch phonology are adult-like at this point. Also looking at spontaneous word productions by 1-year-olds exposed to Tokyo Japanese, Ota (2003b) notes that the contours produced for final-accent and unaccented disyllabic words were mostly flat, even though targetlike productions should show a rise into the second syllable following a phrase-initial intonational pattern (see (7c) and (7e)). These findings indicate that pitch phonology in Japanese is not acquired through unanalyzed learning of contours found in isolated forms (e.g., (7a), (7c), and (7e)). Rather, the contours associated with lexical pitch are learned before, and separately, from the contours associated with phrasal intonation.

The different rates at which lexical pitch accent and phrase-initial rising intonation emerge in production are probably due to several factors. First, pitch accent in Tokyo Japanese is more consistently realized than the phrase-initial rise. Except in phrase-final words with a final accent, a pitch accent specified in the lexicon always manifests itself as a falling contour. In contrast, lowering of a phrase-initial syllable is not attested when the initial syllable is accented or long (i.e., (C)V:, (C)VV or (C)VN) (Pierrehumbert and Beckman 1988). Second, the two types of contour differ in phonetic salience. The range of pitch change in the falling contour of Tokyo Japanese lexical accent is typically larger than that of the rising contour that arises from phrase-initial lowering. Finally, producing the initial rise may be phonetically more challenging than the lexical pitch fall because rising contours require more physiological effort (Snow 1998).

Most of these and other properties of pitch phonology in Tokyo Japanese, including compound accent rules and default accent assignment to loanwords, are acquired by age 5 or 6 (Shirose, Kakehi, and Kiritani 2001; Shirose, Kakehi, and Ōta 2005; Shirose 2009). By this age, children not only understand the global characteristics of Tokyo Japanese pitch phonology and the patterns associated with individual lexical items (i.e., pitch accent), but also some default rules that govern certain groups of words. For example, loanwords in Tokyo Japanese tend to have a typical lexical pitch pattern depending on the syllable structure of the word. Four-syllable words with no heavy syllables are by default unaccented (e.g., *katarogu* 'catalog'). Five-syllable words with no heavy syllables tend to have antepenultimate pitch accent (e.g., *eberésuto* 'Everest'). Heavy syllables attract pitch accent in trisyllabic words (e.g., *tóosuto* 'toast'; *sukánku* 'skunk'). In an elicited production task, Shirose (2009) show that the majority of 5- and 6-year-olds assign these default patterns to made-up words ostensibly presented as country names except in the case of trisyllabic words containing heavy syllables (e.g., *tannoka*, *komonno*), for which the most common response was unaccented. Structures such as *tannoka* and *kommono* consist of 4 moras, just as words with four light syllables (e.g., *notakamo*), which receive an unaccented pattern. It appears, then, 5- to 6-year-olds employ a mora-based analysis in determining the accent pattern of novel foreign-sounding words (Shirose 2009).

4.4 Dialectal differences

Although the preceding discussion of pitch acquisition was based on Tokyo Japanese, there are considerable dialectal differences in the pitch phonology of Japanese (see, among others, Shibatani 1990, Uwano 1999, Kubozono 2012, and various chapters in the Dialect Volume). Accordingly, the developmental pattern differs across dialects. Shirose and her colleagues (Shirose 2007; Shirose, Kakehi, and Kiritani 2002a,b, Shirose, Kakehi, and Ōta 2005) have carried out a series of studies comparing the production of simple words in Tokyo, Kyoto and Kagoshima Japanese, as well as the acceptability judgments of familiar words in Tokyo and Kagoshima Japanese. They consistently found that pitch phonology development in Kagoshima Japanese lagged behind that in Tokyo and Kyoto dialects. In Kagoshima, pitch accent falls on the penultimate or final syllable of the phrase depending on the lexical item that heads the phrase. Words that assign a pitch peak on the phrase's penultimate syllable are called Type A, and those that assign the peak on the phrase-final syllable, Type B. These are illustrated in (8), with the acute accent diacritic indicating the pitch peak. In production, 4-year-olds tend to produce both types of words with a penultimate accent.

(8) Kagoshima pitch accent

Type A	a.	*hána*	'nose'
	b.	*hanáa-ga*	'nose-NOMINATIVE'
Type B	c.	*yamá*	'mountain'
	d.	*yama-gá*	'mountain-NOMINATIVE'

What can account for the developmental pattern in Kagoshima? It is likely that a system such as Kagoshima is more difficult to break into than one like Tokyo or Kyoto. In Tokyo, pitch accent is lexically determined and assigned at the lexical-level, so one only needs to learn the position of the pitch accent for each word. However, in Kagoshima, pitch accent is lexically determined but assigned at the phrase-level. Therefore, the position of the pitch shifts within the word (cf. (8a) vs. (8b) and (8c) vs. (8d)), and yet, they are in part determined by the word. The overgeneralization of the penultimate pattern suggests that children first figure out the assignment domain (i.e., the accentual phrase) of the pitch accent but require some time to understand the status of the assigner (i.e., the two lexical categories). It is not immediately clear why the overgeneralization is made in the direction of the penultimate pattern though. The distribution of pitch types in Kagoshima is not the source of this bias. An examination of words typically occurring in children's lexicon has revealed no statistically significant difference in the frequencies of Type A and Type B words (Shirose, Kakehi, and Kiritani 2002b). Instead, this effect may reflect a general bias toward non-final accent that emerges in early phonological systems. Preference for non-final accentuation has also been reported in children's assignment of compound

accent in Tokyo and Kyoto Japanese (Shirose and Kiritani 2001; Shirose, Kakehi, and Kiritani 2001), which could be the result of a phonetic effect that favors a falling contour coinciding with the end of a prosodic unit (Lieberman 1967).

5 Word-internal prosodic structure

There is converging evidence that a lexical word in Japanese is comprised of the prosodic categories given in Figure 1, organized in a hierarchical structure (Ito 1990; Kubozono 1995, 1999; McCawley 1968; Poser 1990; Vance 1987; but see Labrune 2012 for arguments against positing the syllable for Tokyo Japanese. The Introduction to this volume serves as a useful review of the relevant literature).

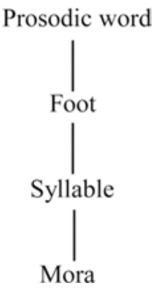


Figure 1: The prosodic hierarchy

This section first reviews the evidence that the same structure can be assigned to the developing phonology of Japanese-speaking children, focusing on the size (section 5.1) and shape (section 5.2) of word production by 1- to 2-year-olds. Section 5.3 deals with when and how children develop the well-documented tendency of Japanese speakers to segment words into units that correspond to the mora.

5.1 Bimoraic minimality in early production

In developmental literature, much interest has been drawn to the extent to which prosodic structural organization exemplified in Figure 1 can be attested in children's phonology (Demuth 1995; Fikkert 1994; Pater 1997). A potential source of evidence for this question is the minimal size of words that children produce. The prosodic hierarchy in Figure 1 is subject to several organizational conditions. A lexical word must be contained in a prosodic word (McCarthy and Prince 1986; Nespor and Vogel 1986). Proper Headedness (Ito and Mester 1992; Selkirk 1996) demands that a prosodic word contains at least one foot, a foot contains at least one syllable, and a syllable, at least one mora. A foot is to be binary, that is, either disyllabic or

bimoraic (Prince 1990). If all these conditions are to be met, a lexical word must contain at least one binary foot, which must have at least two moras; hence a lexical word is minimally bimoraic.

Japanese offers a useful testing ground to examine whether children's early words are subject to a bimoraic lower limit for a couple of reasons. First, the bimoraic minimality condition only applies to derived words in Japanese (Ito 1990), and there are many underived monomoraic words in the language (e.g., *me* 'eye', *te* 'hand', *ha* 'tooth', *ki* 'tree', *e* 'picture'). However, there are reasons to believe that even such words should conform to bimoraic minimality during the earliest stage of development in Japanese. In Optimality Theory, the prosodic conditions mentioned above, such as Proper Headedness and Foot Binarity, are understood as violable *markedness* constraints that are ranked with respect to each other along with *faithfulness* constraints that militate against any modification of the input (Prince and Smolensky 2004). Learnability considerations within this framework (e.g., Smolensky 1996) indicate that in the initial state (i.e., the starting point of all language acquisition) markedness constraints must all be ranked above faithfulness constraints. This means that even in Japanese, any phonological output during the earliest stage of development should fully conform to both Proper Headedness and Foot Binarity, and hence bimoraic minimality. Evidence for this stage should be found in some form of prosodic augmentation in monomoraic words. Second, Japanese does not have stress, and this has an important implication for the interpretation of children's truncated word production. Children frequently omit syllables from the adult word they intend to produce. If there is a bimoraic minimality constraint that regulates early word forms, then such "truncated" output forms should also obey the restriction. In stress languages, such as English and Dutch, however, there is a very strong tendency for children to retain the stressed syllable in truncated productions: e.g., *banána* → [nǽnə], *giráffe* → [waf]. Because stressed syllables in such languages also tend to be heavy (i.e., contain more than one mora), a monosyllabic output form after truncation would automatically be at least two moras long, but this effect cannot be ascribed to a word-size constraint. In contrast, Japanese does not have any syllable weight phenomena tied to intensity-based prominence (i.e., stress), so monosyllabic truncation could, in theory, result in monomoraic forms (e.g., *banana* → [ba]?) unless there is a structurally-motivated lower size limit.

With these observations in the background, let us now turn to some findings that indicate the existence of a bimoraic lower size limit in early Japanese word production. Firstly, although Japanese-speaking children frequently omit syllables in early word production, they exhibit a marked tendency against truncating a target word to monomoraic forms, whether in elicited (T. Itō 2000) or spontaneous production (Ota 2003a). In one estimate, no more than 10% of all monosyllabic truncated forms produced by 1- to 2-year-old children were monomoraic (Ota 2003a). Furthermore, many of the target monomoraic syllables that are retained in the truncation are lengthened in the child's form (Ōkubo 1981; Ingram 1999), and have a falling

pitch contour that can be carried by bimoraic, but not monomoraic, structures in adult Japanese (Ota 2003a). Some examples are given in (9).

(9) Lengthening of monomoraic syllables in truncated forms (Ōkubo 1981; Ota 2003a)

a.	[ita] → [taː]	'there it is'	(T 1;1.16)
b.	[koinobori] → [boː]	'carp streamer'	(T 1;2.0)
c.	[kani] → [kaː]	'crab'	(Hiromi 1;3.4)
d.	[koko] → [koː]	'here'	(Hiromi 1;2.21)
e.	[ana] → [naː]	'hole'	(Kenta 2;2.27)
f.	[banana] → [baː]	'banana'	(Kenta 2;3.22)

A second type of evidence for a bimoraic lower size constraint on early words comes from children's production of monomoraic target words. As with the monomoraic target syllables left in truncated outputs, there is a short period of time during which these words exhibit lengthening (Kawakami and Itō 1999; Ota 2003a). Some examples are given in (10). These productions also exhibit a falling pitch contour that is more comparable to the profile of accented bimoraic words than of monomoirac ones (Ota 2003a). The augmentation cannot be due to a general final lengthening effect because the duration of children's monomoraic target words can be more than twice as long as that of their production of a final CV syllable in CVCV targets (e.g., /mama/ 'mama', Ota 2003a).

(10) Lengthening of monomoraic target words (Ota 2003a)

a.	[me] → [meː] 'eye'	(Hiromi 1;9.11)
b.	[e] → [eː] 'picture'	(Hiromi 1;9.28)
c.	[te] → [teː] 'hand'	(Takeru 1;11.2)
d.	[ʒi] → [diː] 'letter'	(Kenta 2;2.27)
e.	[ki] → [ɟiː] 'tree'	(Kenta 2;2.27)

Older children do not usually truncate words or lengthen monomoraic words to the extent that is described above. However, they show a different type of non-adultlike production pattern that suggests a bimoraic word minimality effect. The phenomenon involves the nominative (*ga*) and dative (*ni*) case marker, which 4- to 5-year-olds occasionally appear to over-apply, as seen in (11).

(11) "Extra case marking" errors by 4-5 year-olds

a. Adult form: *ti* *ga* *de-ta*
blood nom emerge-past
lit. 'I'm bleeding'
Child form: ti ga ga deta

b.	Adult form:	*ka*	*ni*	*sas-are-te*
		mosquito	dat	bite-pass-inf
		lit. 'I got bitten by a mosquito'		
	Child form:	ka ni ni sasarete		

c.	Adult form:	*ka*	*ga*	*i-ta*
		mosquito	nom	be-past
		lit. 'There is a mosquito'		
	Child form:	ka ga ga ita		

d.	Adult form:	*ka*
		mosquito
		lit. 'a mosquito'
	Child form:	ka ga (when shown a picture of a mosquito)

A few additional points need mentioning in order to understand the most plausible explanation of this pattern of errors. First, these types of errors are only attested with monomoraic lexical words. Second, in conversational Japanese, these case markers are optional. Thus, both (12a) and (12b) are possible and attested in child-directed speech.

(12) Optionality of case marking

a.	inu	ga	ita	'There was a dog'
	dog	nom	be-past	

b.	inu	i-ta	'There was a dog'
	dog	be-past	

Third, these errors are confined to words that typically appear in limited syntactic structures. The word *ti* ('blood'), for instance, mostly occurs in the frame … *ga deta*, and *ka* ('mosquito') in … *ni sasareta* or *ga ita*. Given the optionality of case-marking, there are two possible interpretations children can make of a structure such as *ti ga deta* (12a): Either *ti* is the lexical word accompanied by a nominative marker *ga* (the correct analysis) or *tiga* is the lexical word, with the case marker omitted. The errors in (12) indicate that 4-5 year-olds are biased to choose the morphologically wrong but prosodically more harmonic analysis that the lexical word is bimoraic (*tiga*), rather than monomoraic (*ti*).

5.2 Syllable weight effects

An interesting characteristic of the shape of early words in Japanese is the apparent pressure against light-heavy (LH or monomoraic-bimoraic) syllable sequences such

as CVCVV. Two relevant observations are available. First, when longer words are truncated, they result in LH structures much less frequently than in other disyllabic structures, i.e., LL (e.g., [banana] → [bana] 'banana'), HL (e.g., [ʒidoːʃa] → [doːʃa] 'car') and HH (e.g., [ampamman] → [amman] '(name of a cartoon character)' (Ota 1998, 2003a, 2006). Second, while disyllabic words rarely undergo truncation in children's production, the truncation rate for LH words is significantly higher than that for other disyllabic structures (Ota 2006, 2013). Some examples are given in (13). Third, there are reported cases of insertion errors that can be explained as avoidance of LH structures (e.g., [kaban] → [kamban], 'bag,' Fujiwara 1977).

(13) Truncation of LH targets (Ota 2013)
a. [omo'i] → [moi] 'heavy' Ryo (2:0)
b. [kajɯ'i] → [jɯi] 'itchy' Ryo (2;1)
c. [bɯdoː] → [bɯː] 'grape' Tai (1;5)
d. [tokeː] → [keː] 'clock' Tai (1;5-1;6)
e. [suŋo'i] → [goi] 'great' Tai (1;7)

This pattern is mirrored in perceptual experiments, in which 8-10 month old infants show clear preference for HL over LH (Hayashi, Tamekawa, and Mazuka 2000). As pointed out by Kubozono (2000), the anti-LH effect observed here bears striking resemblance to the outputs of adult prosodic morphological operations (e.g., loanword truncation, argot, reduplicative mimetics), which disallow LH. While it is tempting to link the child phenomenon directly to these structural generalizations in adult grammar, there is a more plausible explanation. Research in other languages shows that by the second half of the first year, infants begin to respond to the predominant prosodic pattern of words in their language (Jusczyk, Cutler, and Redanz 1993; Echols, Crowhurst, and Childers 1997; Weber et al. 2004; Friederici, Friedrich, and Christophe 2007). English-exposed infants, for example, begin to exhibit preference for initially-stressed disyllables over finally-stressed disyllables between 6 and 9 months (Jusczyk, Cutler, and Redanz 1993). In the case of Japanese, when one examines all the words used in child-directed speech, LH words are somewhat less frequent than HL, but not so in comparison to HH words (Ota 2006). However, LH structures are conspicuously missing in vocabulary items that are unique to the register used to address infants and young children. The vast majority of such "baby-talk words" are of the shape HL (e.g., *nenne* 'sleep', *anyo* 'foot') or HH (e.g., *ponpon* 'tummy', *tintin* 'penis') (Kubozono 2003). By one estimate, 80% of trimoraic baby-talk words are HL (as opposed to LH or LLL) and 60% of quadrimoraic ones are HH (as opposed to, for example, LLLL or LHL) (Hayashi, Tamekawa, and Mazuka 2000). As such, LH structures may be disfavored by young children because they are noticeably underrepresented among lexical items that are central to their linguistic interaction.

5.3 Mora-based segmentation

Japanese is often called a "mora-timed" language, supposedly belonging to a class of languages different from those that take stress or syllables as timing units (see Otake, this volume, for an overview of this topic). Evidence for moras as isochronous speech units is not particularly robust (Warner and Arai 2001) and attempts to characterize mora-timing based on phonetic measures remain quite elusive (cf. Ramus, Nespor, and Mehler 1999; Grabe and Low 2002; Kohler 2009). However, it is uncontroversial that adult native speakers of Japanese tend to segment words at points that correspond to mora boundaries, often breaking the integrity of the syllable, e.g., *to-o* 'ten', *ho-n* 'book' (Katada 1990; Kubozono 1989, 1995, 1996; Otake et al. 1993). From the perspective of phonological acquisition, this raises the question: When and how does this segmentation pattern develop in Japanese speakers?

A potential source of developmental influence is orthography (Beckman 1995). Of the three writing systems in Japanese, two – hiragana and katakana – are based on moraic units. Each hiragana or katakana symbol corresponds to either a CV unit, the second half of a long vowel or a diphthong, coda nasal, or the first half of a geminate. Therefore, there is a perfect one-to-one correspondence between kana symbols and moras in Japanese writing. Many Japanese children learn the kana writing systems around the time they begin schooling and there is a great deal of evidence that this orthographic exposure is connected to a raised awareness of mora units among children (Mann 1986; Inagaki, Hatano, and Otake 2000). For instance, Inagaki, Hatano, and Otake (2000) tested the segmentation units of 4- to 6-year-olds by administering vocal-motor segmentation tasks (in which the children were asked to make a doll jump on colored circles as they articulate the stimulus word) and a syllable monitoring task (in which the children were asked to detect either a CV unit or CVN unit in the target word). The results from both experiments showed that as children learn how to read kana, their segmentation becomes more mora-based and less syllable-based, except for words containing a geminate consonant.

It is much less clear whether preliterate children also show some tendency toward mora-based segmentation, although such indications were obtained in some studies (Itō and Tatsumi 1997; Kawakami and Itō 1999; Itō and Kagawa 2001). In Itō and Tatsumi (1997), the children (3- to 5-year-olds) were trained to segment words consisting only of monomoraic syllables (e.g., /hasami/ 'scissors' → ha-sa-mi), and then asked to apply the procedure to words containing bimoraic syllables (e.g., /suika/ 'watermelon,' /ringo/ 'apple,' /boosi/ 'hat' and /happa/ 'leaf'). Although only eight of the twenty 4-year-old subjects could read hiragana, all of them separated the bimoraic syllables after the initial CV (i.e., su-i-ka, ri-n-go, bo-o-si), again, except for syllables closed by a geminate.

If preliterate 4-year-old children already have some awareness of mora boundaries, where does this awareness originate? One explanation is that the high frequency of CV syllables in Japanese predisposes the learner to parse speech into units of CVs (Kubozono 1995). Yet another possible source is the various cultural and social activities that promote the notion of moras, including poetry such as haiku and slogans (e.g., a-n-ze-n wa da-se-ru su-pi-i-do da-sa-na-i yu-u-ki 'Safety is about having the courage not to drive at the speed you could make'), and word games such as *shiritori* (Katada 1990; Goetry et al. 2005). It should not escape our attention that the basic operation of *shiritori*, i.e., finding a word that begins with the last mora of the previous target word, does not involve geminates, the one moraic structure that has been shown not to have developed in preliterate children.

6 Morpho-phonological and lexico-phonological phenomena

While the preceding sections focused mostly on the development of phonological patterns in simple words, there are also phonological phenomena that require reference to the morphological composition of complex words and a specific class of words in the language. This section provides developmental discussion of two such cases. The first of these, *rendaku*, is a process that takes place in compounds, and therefore an example of a linguistic pattern that lies at the interface between phonology and morphology. The second relates to the observation that the Japanese lexicon comprises several groups (or sublexica) that have different phonological patterns. As such, it concerns the acquisition of phonology in the broader context of lexical development.

6.1 Rendaku

Rendaku, which induces voicing of the initial obstruent in the second element of a native word compound, has been a focus of intensive research in theoretical phonology (see Vance, this volume, for an overview). Some typical examples are shown in (14). As is well-documented, the process is subject to several constraints, most notably Lyman's Law, a specific form of the Obligatory Contour Principle, which prohibits a morpheme to have more than one voiced obstruent. Lyman's Law blocks rendaku when a voiced obstruent is already in the second member of the compound (14c–d), but not when it is in the first member (14e–f), and it is blind to the non-contrasting voice in sonorants (14g–h).

(14) Rendaku
  a. *kosi* ‘hip’+ *taka* ‘high’→ *kosidaka* ‘reluctance’
  b. *hana* ‘flower’+ *kata* ‘pattern’→ *hanagata* ‘star’
  c. *hana* ‘flower’ + *taba* ‘bunch’ → *hanataba* (**hanadaba*) ‘bouquet’
  d. *ao* ‘blue’ + *kabi* ‘mold’ → *aokabi* (**aogabi*) ‘blue mold’
  e. *hasigo* ‘ladder’ + *sake* ‘sake’ → *hasigozake* ‘bar-hopping’
  f. *geta* ‘wooden clog’+ *kake* ‘wear’ → *getagake* ‘wearing of wooden clogs’
  g. *sake* ‘sake’ + *taru* ‘barrel’ → *sakedaru* ‘barrel for sake’
  h. *hira* ‘even’ + *kana* ‘kana symbols’ → *hiragana* ‘hiragana symbols’

Rendaku poses some interesting questions from the perspective of language learning. First, when do children learn the phonological process that underlies the alternation? We know from better-studied patterns, such as the development of the English regular plural (e.g., dog[z], cat[s] and fox[əz]), that frequently attested morphologically complex forms can be learned as unanalyzed units without acquiring the phonological alternation. This may also be the case for rendaku. Therefore, the acquisition of rendaku can only be revealed in productive application to novel compounds. Second, what paths do children follow in learning the conditions of rendaku application? Over-application to cases like (14c–d) and under-application to cases like (14e–h) uncovers the nature of phonological generalizations children make when they are exposed to evidence of these alternations.

With respect to the first question, elicited production tasks administered by Fukuda (2002) and Fukuda and Fukuda (1999) show that children before the age of 5½ years have limited overall application of rendaku (below 35% of the test stimuli) even in frequent compounds. Slightly older children (5½ – to 6-year-olds) display a high rate of application for attested compounds but still fail to extend rendaku reliably to novel stimuli. This asymmetry disappears in children older than 6. These results indicate that the productive application of rendaku is typically acquired some time between the ages of 5 and 6 years.

Exploring the second question, Fukuda (2002) compared children’s application of rendaku in contexts corresponding to examples (14a–b), where it applies; (14c–d), where it is blocked by Lyman’s Law; and (4e–f), where it is not blocked by the presence of the voiced obstruent in the first element of the compound. Virtually no cases of over-application to the Lyman’s Law context (as in (14c–d)) were observed. As mentioned above, children under the age of 5½ years often failed to apply rendaku in legitimate contexts, but there was a slight tendency for them to apply rendaku more readily when the first element of the compound lacked a voiced obstruent (as in (14a–b)) than when it contained one (as in (14e–f)).

The study distinguished three types of words depending on their frequencies (frequent vs. non-frequent vs. novel) and three phonological contexts (no voiced obstruent in the first element (N1) vs. voiced obstruent in the second element (N2)

vs. voiced obstruent in N1). Because the analysis does not fully cross these two dimensions, it is not immediately clear whether we can rule out the possibility that the effect was due to the children simply reproducing the voicing in familiar compounds in the frequent category. If, however, the effect was generalizable across the frequency categories, it suggests that children at this stage are beginning to learn the role of contrastive voicing in rendaku, but have not quite identified the relevant domain in which another voiced obstruent blocks sequential voicing. Taking the latter interpretation, Fukuda (2002) attributes the blocking of rendaku by voiced obstruents in N1 to an intermediate ranking of two Optimality Theoretic constraints: ANCHOR[vce], a constraint that induces rendaku (or the alignment of a featural affix [voice] to the left edge of a morphological head), and OCP(vce)$_{\mathrm{PWd}}$, a constraint manifesting a type of the Obligatory Contour Principle that militates against two contrastive voice features within a prosodic word (i.e., the compound as a whole, in this context). The key idea here is that such constraints are present in any phonological grammar, developing or otherwise, whether or not their effects are immediately visible in the language. The learner's task is to determine how these constraints are ranked with respect to each other and other constraints. In the adult state of Japanese, these constraints must be ranked in the order ANCHOR[vce] » OCP(vce)$_{\mathrm{PWd}}$, where ANCHOR[vce] overrides the effects of OCP(vce)$_{\mathrm{PWd}}$ because the presence of a contrastively voiced consonant within *the same compound* does not block rendaku (it is the presence of a voiced obstruent in *the same component word within the compound* that blocks it). But during development when learners are still sorting out the adult ranking, they may go through a stage where the ranking relationship is reversed (OCP(vce)$_{\mathrm{PWd}}$ » ANCHOR[vce]), resulting in a state where rendaku is blocked when it leads to any two contrastive voice features in the compound.

6.2 The phonological lexicon

As with many other languages, Japanese has a lexicon that contains subsets of vocabulary items with different phonological characteristics (for an overview, see Nasu, this volume, on mimetic phonology, Ito and Mester, this volume, on Sino-Japanese phonology, and Kubozono, this volume, on loanword phonology). Much of this compartmentalization of the phonological lexicon in Japanese is due to major influxes of loanwords over the history of the language, which resulted in three commonly identified sublexica: native words (*wago* or *Yamato-kotoba*), Sino-Japanese words (*kango*, which are borrowings from Chinese), and "foreign" words (or *gairaigo*, borrowings from other sources, such as Dutch, Portuguese, and English). The literature often lists a fourth sublexicon known as mimetics and distinguishes two groups of words among the "foreign" category: "assimilated foreign words" and "unassimilated foreign words".

In an influential series of papers on the phonological lexicon of Japanese, Ito and Mester (1995a,b, 1999) made an astute observation that such subsets of lexical items are best characterized in terms of layers of phonological generalizations that apply to one, some, or all groups of words in the language. Thus, the prohibition of post-nasal voiceless obstruents is observed only among native words, the prohibition of initial /p/ is observed among native and Sino-Japanese words, and the prohibition of voiced geminate is observed in all except unassimilated foreign words. The sublexica are Cartesian products of the sets defined by these phonological properties. Crucially, these layered generalizations apply both to alternation and distribution patterns. In the native sublexicon, for example, the ban on postnasal voiceless obstruents triggers voicing alternations like the ones in (15a), while it also accounts for the lack of native morphemes that contain such sequences (e.g., **tonpo*, **kankae*). The alternation and distribution patterns in Sino-Japanese words in (16) shows that this generalization does not apply to other sublexica.

(15) Native sublexicon
  a. Post-nasal voicing
    mi +*ta* → *mita* 'see-PAST'
    tabe + *ta* → *tabeta* 'eat- PAST'
    sin + *ta* → *sinda* 'die- PAST'
    nom + *ta* → *nonda* 'drink- PAST'
  b. *tonbo, kangae, tonbi*

(16) Sino-Japanese sublexicon
  a. Lack of post-nasal voicing
    a. *si* + *teki* → *siteki* 'personal'
    b. *but* + *teki* → *butteki* 'material'
    c. *sin* + *teki* → *sinteki* 'mental'
    d. *tan* + *teki* → *tanteki* 'straightforward'
  b. *kantan, kandoo, hontoo*

The learnability question that arises from this state of affairs is how the child, who encounters input data that consist of the union of (15) and (16), learns that different phonological generalizations apply to (15) and (16) without a priori knowledge of the items' memberships to separate groups. On hearing words from non-native sublexica such as *kantan* 'easy', *ponpon* 'tummy', *hanbun* 'half', and *panda* 'panda', the child is likely to conclude that voicing is generally contrastive for post-nasal obstruents. After that, no positive distributional evidence seems capable of reversing that conclusion for a subset of lexical items in the language (e.g., *tonbo*).

One extreme reaction to this grim scenario is to reject the notion of lexical stratification altogether (Rice 1997). This move does solve the learnability problem by removing the explanandum, but it comes with the expensive cost of losing the most obvious explanations for many systematic and productive alternation patterns such as (15a). To argue that the child cannot learn some phonological patterns that apply to a subset of the lexicon amounts to saying that the mechanism that generates alternation patterns like (15a) belongs outside the realm of phonological grammar, a conclusion that is hard to accept. Less radically, Ota (2004) concurs with Rice (1997) in that lexical stratification is unlearnable based solely on distributional evidence, but argues that language-internal inconsistencies in phonological generalizations can be learned from alternation evidence. Learners exposed to (15) and (16) may arrive at the superset phonological grammar in which voicing is generally contrastive, but are still able to retract that generalization for words that undergo alternation patterns such as (15a). The prediction that follows from this claim is that learners should have no reason to classify two words, such as the native *tombo* and foreign *kombo*, into separate sublexica when there is no surface distributional evidence that compels them to do so. However, there is psycholinguistic evidence that adult speakers of Japanese do treat words from putative sublexica differently even in the absence of direct alternation evidence (Gelbart and Kawahara 2007; Moreton and Amano 1999).

Another solution, couched in Optimality Theory, is to appeal to ranking conservatism (Ito and Mester 1999; Pater 2005). From an initial state in which all markedness constraints are ranked above faithfulness constraints (a widely held hypothesis due to Smolensky 1996), the learner is seen to promote a faithfulness constraint above the relevant markedness constraint only when there is evidence to do so for a particular lexical item. On this account, a child learning Japanese should reverse the initial ranking relationship *NT (no postnasal voiceless obstruents) » FAITH for items such as *kantan* and *ponpon* so that a voiceless obstruent following a nasal is allowed to surface faithfully, but for lack of any specific evidence, the child maintains the initial ranking for *tonbo* and *kangae*, as well as *mita* and *sinda*, where the ranking derives the alternations. One problem with this proposal is that it leads the learner to group the lexical items in a way that does not accord with the typical division of sublexica. For instance, the Sino-Japanese *zenbu* ('all') and the 'foreign' *bando* ('band') would be grouped with the native *tonbo* under *NT » FAITH, against our intuition that the former two should be under FAITH » *NT where voicing contrast exists. Rice (1997) offers a vivid reductio ad absurdum when she points out that applied indiscriminately, this strategy would lead a learner to the conclusion that *fond* and *font* belong to different sublexica in English, one that obeys *NT and one that does not.

While there may be other computational proposals that produce the correct results using purely phonological data, one should also be ready to entertain the possibility that the acquisition of lexical stratification is aided by extraphonological

or even extralinguistic information (Gelbart and Kawahara 2007; Ota 2010). Sublexica in Japanese have skewed distribution in syntactic categories, with some having access to a range of categories (e.g., native words are found in nouns, verbs, adjectives, adverbs, function words etc.) while others restricted to certain categories (e.g., foreign words are generally confined to nouns and light-verb constructions; mimetics to nouns and adverbs). Lexical stratification also has consequences for morphological combinations. For example, there is a strong tendency in Japanese compounds to extract morphemes from the same sublexicon (e.g., *asakusadera* < asa 'morning (native)' + kusa 'grass (native)' + tera 'temple (native)' versus *sensoozi* < sen 'morning (SJ)' + soo 'grass (SJ)' + zi 'temple (SJ)'). Of course, such grammatical information is of no value to learners who have not learned the sublexical division, but it assists in assigning membership to words and morphemes that otherwise may not be classifiable solely on phonological distributional evidence. Furthermore, sublexical division in Japanese is conspicuously mapped onto orthographic conventions by which Sino-Japanese words are typically written in Chinese characters and "foreign" words in katakana. Speakers' intuition about sublexicon membership may also be enforced by such orthographic knowledge.

7 Conclusions and future directions

It should be clear from this short survey that a great deal of research has been carried out in the past decades on the acquisition of Japanese phonology. This effort covers many aspects of the sound system that are of central interest to the phonology of Japanese itself, such as durational contrasts, pitch accent, mora-based segmentation, rendaku, and the stratification of the lexicon. Developmental research in this area has contributed to our understanding of these phonological properties. For example, the close link between kana literacy acquisition and the development of mora-based segmentation suggests that the adult speakers' intuition on this phenomenon is likely related to orthographic knowledge. Formal learnability issues in the acquisition of lexical stratification have honed our theorizing of the relationship between distribution, alternations, and input data for grammar constructions.

Because much work in modern phonological acquisition is based on the learning of English and other European languages, research on Japanese phonological development has helped counterbalance this empirical bias by bringing typologically different phenomena to the table. The acquisition of lexical pitch accent offers insights into the development of a prosodic system that are different from what can be gained through studying the development of lexical stress, for example, as the lexical property of the prosody shares the phonetic space (i.e., pitch) with the non-lexical component of the prosodic system (i.e., intonation). The abundance of monomoraic words in Japanese presents a testing ground for the putative bimoraic

minimality effect that cannot be examined in the development of languages such as English and German, in which all lexical words in the input are already minimally bimoraic. The combination of long-distant process, abstract feature blocking and structural constraints that governs rendaku provides a unique empirical domain in which we can examine how the interaction of such factors may figure in the acquisition of the phonology-morphology interface.

Cross-linguistic comparisons between Japanese and other languages have also advanced our understanding of the mechanisms and factors behind child phonology phenomena and developmental processes. Similarities and differences found in segmental acquisition order and substitution errors can disentangle the effects of general articulatory demands, phonetic details of otherwise comparable target segments, proximity and similarity to other segments in the inventory, and the frequency distribution of the relevant sounds in the input. Comparison of the primarily spectrally-cued vowel contrasts in English and the durationally-cued vowel contrasts in Japanese has allowed us to examine the plausibility of distributional learning in the context of multi-dimensional phonetic space.

While these are notable achievements, it goes without saying that much remains to be learned about the acquisition of Japanese phonology. Despite the substantial body of previous research on children's segmental production, our understanding of how segmental development unfolds in Japanese and what exactly causes the documented divergence in early production is still fairly limited. The emergence of infant perception studies in this area and more phonetically sophisticated examination of production data (along the lines of Li, Edwards, and Beckman 2009) should go far toward meeting this demand. Other areas that could benefit from more empirical work include the development of the functional roles of intonation (although see Ichijima 2009 and Ito et al. 2012), pitch accent patterns in morphologically complex forms (e.g., noun phrases, verb phrases, complex compounds; see Shirose and Kiritani 2001 for simple compounds), and vowel devoicing in early ages.

Much of the recent work carried out on Japanese follows the standard methods used in mainstream research on the development of speech, phonetics and phonology, as well as the strategies in studying a single language or a combination of typologically different languages. However, one approach that has proven particularly fruitful in the study of Japanese phonological acquisition is cross-dialectal comparison, where the developmental patterns of two or more dialects with some pertinent differences are examined in order to understand the relative contributions of the linguistic factors or the mechanisms that underlie the learning processes. Already successfully applied to pitch phonology (Shirose, Kakehi, and Ōta 2005) and vowel devoicing (Imaizumi, Fuwa, and Hosoi 1999), it is likely that this approach can yield similarly interesting results in other areas of the development of Japanese phonology such as velar nasalization and the relative roles of the syllable and mora in prosodic structures.

References

Akamatsu, Tsutomu. 1997. *Japanese phonetics: Theory and practice*. Munich: Lincom Europa.

Aoyama, Katsura. 2001. *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions*. Boston, MA: Kluwer Academic Publishers.

Baum, Shari R. and James C. McNutt. 1990. An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics* 18(1). 51–63.

Best, Catherine C. and Gerald W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech* 46(2–3). 183–216.

Best, Catherine T., Gerald W. McRoberts, Rosemarie LaFleur and Jean Silver-Isenstadt. 1995. Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development* 18(3). 339–350.

Beckman, Mary. 1995. On blending and the mora. In Bruce Connell and Amalia Arvanit (eds.), *Papers in laboratory phonology IV*, 157–167.

Beckman, Mary E., Kiyoko Yoneyama and Jan Edwards. 2003. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7. 18–28.

Bernhardt, Barbara Handford and Joseph Paul Stemberger. 1998. *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

Boysson-Bardies, Bénédicte de and Marilyn May Vihman. 1991. Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67(2). 297–319.

Brown, Cynthia and John Matthews. 1997. The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs and Martha Young-Scholten (eds.), *Focus on phonological acquisition*, 67–112. Amsterdam: John Benjamins.

Chevrie-Muller, C. and M. T. Lebreton. 1973. Étude de la réalization des consonnes au cours d'une épreuve de mots, sur des groupes d'enfants de 3 ans et 5 ans ½. *Revue de laryngologie* 94. 109–152.

Demuth, Katherine. 1995. Markedness and the development of prosodic structure. *Proceedings of the North East Linguistic Society 25*, 13–26. Amherst, MA: GLSA.

Echols, Catharine H., Megan J. Crowhurst and Jane B. Childers. 1997. The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language* 36(2). 202–225.

Edwards, Jan and Mary E. Beckman. 2008a. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development* 4(2). 122–156.

Edwards, Jan and Mary E. Beckman. 2008b. Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics* 22(12). 937–956.

Fikkert, Paula. 1994. *On the acquisition of prosodic structure*. Dordrecht, Holland: Institute of Generative Linguistics.

Fletcher, Samuel G. and Dennis G. Newman. 1991. [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *Journal of the Acoustical Society of America* 89(2). 850–858.

Friederici, Angela D., Manuela Friedrich and Anne Christophe. 2007. Brain responses in 4-month-old infants are already language specific. *Current Biology* 17(14). 1208–1211.

Fujimoto, Masako. this volume. Chapter 4: Vowel devoicing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Fujiwara, Yōichi. 1977. *Yōji no gengo hyōgen nōryoku no hattatsu* [The development of children's language production skills]. Hiroshima: Bunka Hyōron Shuppan.

Fukuda, Suzy E. 2002. The acquisition of rendaku: A linguistic investigation. *Aoyama Journal of General Education* 43. 15–59.

Fukuda, Suzy E. and Shinji Fukuda. 1999. The operation of rendaku in the Japanese specifically language-impaired: A preliminary investigation. *Folia Phoniatrica et Logopaedica* 51(1–2). 36–54.

Gelbart, Ben and Shigeto Kawahara. 2007. Lexical cues to foreignness in Japanese. *Proceedings of Formal Approaches to Japanese Linguistics 4*, 55–67.

Goetry, Vincent, Satsuky Urbain, José Morais and Régine Kolinsky. 2005. Paths to phonemic awareness in Japanese: Evidence from a training study. *Applied Psycholinguistics* 26(2). 285–309.

Goldstein, Brian A., Leah Fabiano and Patricia Swasey Washington. 2005. Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools* 36(3). 201–218.

Grabe, Esther and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 515–546.

Hallé, Pierre A., Bénédicte de Boysson-Bardies and Marilyn M. Vihman. 1991. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34(4). 299–318.

Han, Mieko S. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49(2). 102–127.

Hayashi, Akiko, Yūji Tamekawa and Reiko Mazuka. 2000. Muimigo no rizumu patān ni kansuru nyūji no kanjusei no hattatsu [The development of children's sensitivities to rhythmic patterns in nonsense words]. *Bulletin of the Research Institute for the Education of Exceptional Children*. 67–74. Tokyo Gakugei University.

Hirata, Yukari. this volume. Chapter 18: L2 phonetics and phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Hirata, Yukari and Kimiko Tsukada. 2009. Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* 66(3). 129–149.

Hoequist, Charles. 1983a. Durational correlates of linguistic rhythm categories. *Phonetica* 40(1). 19–31.

Hoequist, Charles. 1983b. Syllable duration in stress-, syllable-, and mora-timed languages. *Phonetica* 40(3). 203–237.

Ichijima, Tamiko. 2009. Nihongo ni okeru shoki onritsu to on'in no kakutoku [Early prosody and phonological acquisition in Japanese]. *Humanity and Science: Journal of the Faculty of Health and Welfare* 9(1). 15–28.

Igarashi, Yosuke. this volume. Chapter 13: Intonation. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Imaizumi, Satoshi, Kiyoko Fuwa and Hiroshi Hosoi. 1999. Development of adaptive phonetic gestures in children: Evidence from vowel devoicing in two different dialects of Japanese. *Journal of the Acoustical Society of America* 106(2). 1033–1044.

Inagaki, Kayoko, Giyoo Hatano and Takashi Otake. 2000. The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology* 75(1). 70–91.

Ingram, David. 1974. Phonological rules in young children. *Journal of Child Language* 1(1). 49–64.

Ingram, David. 1989. First language acquisition: Method, description and explanation. Cambridge: Cambridge University Press.

Ingram, David. 1999. The nature of prosodic primitives: Evidence from child language and language impairment. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 34-II*, 269–276.

Ito, Junko. 1990. Prosodic minimality in Japanese. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II*, 213–239.

Ito, Junko and Armin Mester. 1992. *Weak layering and word binarity* (LRC-92-09). Santa Cruz, CA: Linguistic Research Center, University of California, Santa Cruz.

Ito, Junko, and Armin Mester. 1995a. The core-periphery structure of the lexicon and constraints on reranking. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers in Linguistics 18), 181–209. Amherst: GLSA.

Ito, Junko, and Armin Mester. 1995b. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Ito, Junko and Armin Mester. this volume. Chapter 7: Sino-Japanese phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Itō, Katsutoshi. 1990. *Kodomo no kotoba* [Children's language]. Tokyo: Keisō Shobō.

Ito, Kiwako, Nobuyuki Jincho, Utako Minai, Naoto Yamane and Reiko Mazuka. 2012. Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language* 66(1). 265–284.

Itō, Tomohiko. 2000. Yōji no onseikei no kanryakuka genshō to inritsu tan'i no kakutoku [Simplification phenomena in children's phonetic forms and the acquisition of prosodic segmentation units]. *Proceedings of the Japan Society of Logopedics and Phoniatrics 45*, 53.

Itō, Tomohiko and Aya Kagawa. 2001. Moji kakutoku mae no yōji ni okeru onritsu tan'i no hattatsu [The development of prosodic units in preliterate children]. *Onsei Gengo Igaku* 42(3). 235–241.

Itō, Tomohiko and Itaru Tatsumi. 1997. Tokushu haku ni taisuru metagengo chishiki no hattatsu [The development of metalinguistic knowledge on special moras]. *Onsei Gengo Igaku* 38. 196–203.

Jusczyk, Peter W., Anne Cutler and Nancy J. Redanz. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64(3). 675–687.

Kajikawa, Sachiyo, Laurel Fais, Ryoko Mugitani, Janet F. Werker and Shigeaki Amano. 2006. Cross-language sensitivity to phonotactic patterns in infants. *Journal of the Acoustical Society of America* 120(4). 2278–2284.

Karzon, Roanne Gottlieb. 1985. Discrimination of polysyllabic sequences by one-to four-month-old infants. *Journal of Experimental Child Psychology* 39(2). 326–342.

Karzon, Roanne Gottlieb and Johanna G. Nicholas. 1989. Syllabic pitch perception in 2-to 3-month-old infants. *Attention, Perception, & Psychophysics* 45(1). 10–14.

Katada, Fusa. 1990. On the representation of moras: Evidence from a language game. *Linguistic Inquiry* 21(4). 641–646.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.

Kawakami, Shin. 1977. *Nihongo onsei gaisetsu* [Introduction to the sounds of Japanese]. Tokyo: Ōfūsha.

Kawakami, Masayo and Tomohiko Itō. 1999. Yōji no tango koshō ni okeru hiryūchōsei no shutsugen to hatsuwa tan'i no kakutoku [Occurrences of disfluency in children's word naming and the acquisition of speech units]. *Tokushu Kyōikugaku Kenkyū* 37(2). 23–30.

Kent, Ray. 1992. The biology of phonological development. In Charles Ferguson, Lise Menn and Carol Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, 65–90. Timonium, MD: York Press.

Kohler, Klaus J. 2009. Rhythm in speech and language. *Phonetica* 66(1–2). 29–45.
Kubozono, Haruo. 1989. The mora and syllable structure in Japanese. *Language and Speech* 32(3). 249–278.
Kubozono, Haruo. 1995. *Gokeisei to on'in kōzō* [Word Formation and phonological Structure]. Tokyo: Kurosio.
Kubozono, Haruo. 1996. Speech segmentation and phonological structure. In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*, 77–94, Berlin: Mouton de Gruyter.
Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.
Kubozono, Haruo. 2000. Nihongo no goi to onsetsu kōzō [Japanese vocabulary and syllable structure]. *Nihongo Kenkyū* 20. 1–18.
Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In Caroline Féry and Ruben van de Vijver (eds.), *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press.
Kubozono, Haruo. 2012. Varieties of pitch accent systems in Japanese. *Lingua* 122. 1395–1414.
Kubozono, Haruo. this volume. Chapter 8: Loanword phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Kuhl, Patricia K. and James D. Miller. 1982. Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Attention, Perception, & Psychophysics* 31(3). 279–292.
Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani and Paul Iverson. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9(2). F13–F21.
Kunnari, Sari, Satsuki Nakai and Marilyn M. Vihman. 2001. Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication* 5(2). 13–24.
Labrune, Laurence. 2012. Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29(1). 113–152.
Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics.* Chicago: Chicago University Press.
Li, Fangfang, Jan Edwards and Mary E. Beckman. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics* 37(1). 111–124.
Lieberman, Philip. 1967. Intonation, perception, and language. Cambridge, MA: MIT Press.
Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.
Macken, Marlys A. and David Barton. 1980. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7(1). 41–74.
Mann, Virginia A. 1986. Phonological awareness: The role of reading experience. *Cognition* 24(1). 65–92.
Masuko, Yukie. 2004. Nihongo no gengo-kakutoku-katei ni okeru onsei hyōshutsu ni kansuru kenkyū [A study on the sounds produced by an infant in the process of acquiring Japanese]. *Tokyo Gaikokugo Daigaku Ronshū* 67. 15–37.
Maxwell, Edith M. and Gary Weismer. 1982. The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics* 3(1). 29–43.
Maye, Jessica, Janet F. Werker and LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3). B101–B111.
Maye, Jessica, Daniel J. Weiss and Richard N. Aslin. 2007. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science* 11(1). 122–134.

Mazuka, Reiko. 2010. Learning the sound system of Japanese: What does it tell us about language acquisition? *Proceedings of the International Congress on Acoustics 20*, 4186–4193.

McCarthy, John J. and Alan Prince. 1986. *Prosodic morphology* (RuCCs Technical Report Series TR-3). New Brunswick, NJ: Rutgers Center for Cognitive Science.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Minagawa-Kawai, Yasuyo, Koichi Mori, Nozomi Naoi and Shozo Kojima. 2007. Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *The Journal of Neuroscience* 27(2). 315–321.

Moreton, Eliot and Shigeaki Amano. 1999. The effect of lexical stratum phonotactics on the perception of Japanese vowel length. *IEIC Technical Report* (Institute of Electronics, Information and Communication Engineers) 99(166). 1–8.

Mugitani, Ryoko, Laurel Fais, Sachiyo Kajikawa, Janet F. Werker and Shigeaki Amano. 2007. Age-related changes in sensitivity to native phonotactics in Japanese infants. *Journal of the Acoustical Society of America* 122(3). 1332–1335.

Mugitani, Ryoko, Ferran Pons, Laurel Fais, Christiane Dietrich, Janet F. Werker and Shigeaki Amano. 2009. Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology* 45(1). 236–247.

Murata, Kōji. 1970. *Yōji no kotoba to hatsuon* [Children's language and articulation]. Tokyo: Baifūkan.

Nakajima, Makoto, Jun'ichi Murai, Shigeo Okuno, Masumi Shimizu, Natsuki Okamoto, Masato Tanaka and Tamiko Maeda. 1962. Onsei no kigōka narabi ni taisei katei ni kansuru kenkyū [A study on the symbolization and systematization of phonetics]. *Shinrigaku Hyōron* 6. 1–48.

Nakanishi, Yasuko. 1982. Kōon hattatsu [Articulatory development]. In Hiroshi Uchisugawa and Taiko Nagasawa (eds.), *Kōza gengo shōgai chiryō kyōiku 4: Kōon shōgai* [Series in language disorder therapy education 4: Articulatory disorders], 37–64. Tokyo: Fukumura Shuppan.

Nakanishi, Yasuko, Kenjirō Ōwada and Noriko Fujita. 1972. Kōon kensa to sono kekka ni kansuru kōsatsu [Discussion of articulatory testing and its results]. *Tokyo Gakugei University Bulletin of the Research Institute for the Education of Exceptional Children* 1. 1–52.

Nasu, Akio. this volume. Chapter 6: The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Nazzi, Thierry, Caroline Floccia and Josiane Bertoncini. 1998. Discrimination of pitch contours by neonates. *Infant Behavior and Development* 21(4). 779–784.

Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Noda, Masako, Yumiko Iwamura, Keiko Naitō and Kiyomi Asukai. 1969. Yōji no kōon nōryoku no hattatsu ni kansuru kenkyū [A study on the development of children's articulatory ability]. *Nihon Sōgō Aiiku Kenkyūjo Kiyō* 4. 153–170.

Okada, Hideo. 1999. Japanese. In the International Phonetic Association (ed.), *Handbook of the International Phonetic Association*, 117–119. Cambridge: Cambridge University Press.

Ōkubo, Ai. 1977. *Yōji no kotoba to otona* [Children's language and adults]. Tokyo: Sanseido.

Ōkubo, Ai. 1981. *Yōji no kotoba shiryō* [Data of children's language]. Tokyo: Shūei Shuppan.

Ota, Mitsuhiko. 1998. Minimality constraints and the prosodic structure of child Japanese. In David Silva (ed.), *Japanese/Korean Linguistics 8*, 331–344. Stanford: CSLI.

Ota, Mitsuhiko. 2003a. *The development of prosodic structure in early words: Continuity, divergence, and change*. Amsterdam: John Benjamins.

Ota, Mitsuhiko. 2003b. The development of lexical pitch accent systems: An autosegmental analysis. *Canadian Journal of Linguistics* 38(2). 357–383.

Ota, Mitsuhiko. 2004. The learnability of the stratified phonological lexicon. *Journal of Japanese Linguistics* 20(4). 19–40.

Ota, Mitsuhiko. 2006. Input frequency and truncation in child Japanese: Structural and lexical effects. *Language and Speech* 49(2). 261–295.
Ota, Mitsuhiko. 2010. Kango on'in no shūtoku kanōsei ni tsuite [The learnability of Sino-Japanese phonology]. In Hiroko Oshima, Akiko Nakajima and Raul Blin (eds.), *Kango no gengogaku*, 39–55. Tokyo: Kurosio.
Ota, Mitsuhiko. 2013. Lexical frequency effects on phonological development: The case of word production in Japanese. In Marilyn M. Vihman and Tamar Keren-Portnoy (eds.), *Child Phonology: Whole word approaches, cross-linguistic evidence*, 415–438. Cambridge: Cambridge University Press.
Ota, Mitsuhiko and Isao Ueda. 2006. Speech acquisition of Japanese. In Sharryne McLeod (ed.), *The international guide to speech acquisition*, 457–471. Clifton Park, NY: Delmar Thompson.
Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology.* Berlin: De Gruyter Mouton.
Otake, Takashi, Anne Cutler, Giyoo Hatano and Jacque Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32(2). 258–278.
Ōwada, Kenjirō and Yasuko Nakanishi. 1971. Yōji no kotoba no hattatsu (6) – Kōon no kōjōsei to kansei jun'i [Children's language development (6) – On the stability and acquisition order of articulation]. *Jikō* 43(3). 183–192.
Ōwada, Kenjirō, Yasuko Nakanishi and Katsutoshi Ōshige. 1969. Hoikuenji no kōon no henka ni tsuite [On the articulatory changes in preschoolers]. *Jikō* 41(4). 227–231.
Pater, Joe. 1997. Minimal violation and phonological development. *Language Acquisition* 6(3). 201–253.
Pater, Joe. 2005. Learning a stratified grammar. *Proceedings of the 29th Boston University Conference on Language Development*, 482–492.
Pierrehumbert, Janet and Mary Beckman. 1988. *Japanese tone structure*. Cambridge: MIT Press.
Poser, William J. 1990. Evidence for foot structure in Japanese. *Language* 66(1). 78–105.
Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Proceedings of the Annual Meeting of the Chicago Linguistic Society 26-II*, 355–398.
Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Oxford: Blackwell.
Ramus, Franck, Marina Nespor and Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265–292.
Rice, Keren. 1997. Japanese NC clusters and the redundancy of postnasal voicing. *Linguistic Inquiry* 28(3). 541–551.
Rice, Keren and Peter Avery. 1995. Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In John Archibald (ed.), *Phonological acquisition and phonological theory*, 23–42. Hillsdale, NJ: Lawrence Erlbaum.
Sakauchi, Tomoko. 1967. Kodomo no kōon nōryoku ni tsuite [On children's articulatory ability]. *Gengo Shōgai Kenkyū* 68. 13–26.
Sato, Yutaka, Mahoko Kato and Reiko Mazuka. 2012. Development of single/geminate obstruent discrimination by Japanese infants: Early integration of durational and nondurational cues. *Developmental Psychology* 48(1). 18–34.
Sato, Yutaka, Yuko Sogabe and Reiko Mazuka. 2010a. Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology* 46(1). 106–119.
Sato, Yutaka, Yuko Sogabe and Reiko Mazuka. 2010b. Development of hemispheric specialization for lexical pitch–accent in Japanese infants. *Journal of Cognitive Neuroscience* 22(11). 2503–2513.
Selkirk, Elizabeth O. 1996. The prosodic structure of function words. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
Shirose, Ayako. 2007. Yōji no tango akusento no chōshu ni kansuru hōgen hikaku ni yoru kentō [A comparative study of the perception of Japanese lexical accent in children]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 11(3). 55–68.
Shirose, Ayako. 2009. 4, 5 hakugo no gairaigo akusento kisoku to yōji no hatsuwa seisei [Accent rules in 4-and 5-syllable loanwords and children's speech production]. *Tokyo Gakugei Daigaku Kiyō* 60. 177–194.
Shirose, Ayako, Kazuhiko Kakehi and Shigeru Kiritani. 2001. Interference of accent system variation with the acquisition of compound accent rule: A cross-dialectal study. *Journal of Cognitive Science* 2(1). 33–44.
Shirose, Ayako, Kazuhiko Kakehi and Shigeru Kiritani. 2002a. Kagoshima hōgen yōji no tango akusento kakutoku ni kansuru yobiteki kentō [A preliminary examination of children's accent acquisition of the Kagoshima dialect]. Paper presented at the annual meeting of the Acoustical Society of Japan, Tokyo Joshi Daigaku, 28–29 September.
Shirose, Ayako, Kazuhiko Kakehi and Shigeru Kiritani. 2002b. Akusento no kakutoku to yōji goi ni okeru hōgen akusento no seiki bunpu [Accent acquisition and frequency distribution of dialectal accent in children's vocabulary]. Paper presented at the annual meeting of the Acoustical Society of Japan, Kanagawa Univeristy, 18–20 March.
Shirose, Ayako and Shigeru Kiritani. 2001. Fukugō meishi akusento kisoku no kakutoku katei [Acquisition of compound accent rule in Japanese]. *Onsei Kenkyū* [Journal of the Phonetic Society of Japan] 5(2). 39–53.
Shirose, Ayako, Kazuhiko Kakehi and Ichiro Ōta. 2005. Akusento no kakutoku katei no gengokan hikaku [A crosslinguistic comparison of accent acquisition]. *Denshi Jōhō Tsūshin Gakkai Hōkoku* 105–437. 13–18.
Smit, Ann Bosma. 1993. Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: consonant singletons. *Journal of Speech, Language and Hearing Research* 36(3). 533–547.
Smit, Ann Bosma, Linda Hand, J. Joseph Freilinger, John E. Bernthal and Ann Bird. 1990. The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55(4). 779–798.
Smith, Neilson V. 1973. *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
Smolensky, Paul. 1996. *The initial state and the 'richness of the base' in Optimality Theory* (Rutgers Optimality Archive [ROA-154-1196]). Baltimore, MD: Johns Hopkins University.
Snow, David. 1998. Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language and Hearing Research* 41(3). 576–587.
Stone, Maureen, Alice Faber, Lawrence J. Raphael and Thomas H. Shawker. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics* 20(2). 253–270.
Takagi, Shun'ichirō and Akiko Yasuda. 1967. Seijō yōji (3 – 6 sai) no kōon nōryoku [The articulatory ability of typically developing children (ages 3 to 6)]. *Shōni Hoken Kenkyū* 25. 23–28.
Ueda, Isao. 1996. Segmental acquisition and feature specification in Japanese. In Barbara Bernhardt, John Gilbert and David Ingram (eds.), *Proceedings of the UBS International Conference on Phonological Acquisition*, 15–27. Somerville, MA: Cascadilla Press.
Ueda, Isao, Tomohiko Ito and Tomohiko Shirahata. 1998. A paradox in process-based analysis. In Wolfram Ziegler and Karin Deger (eds.), *Clinical phonetics and linguistics*, 123–130. San Diego: Singular Publishing Company.
Umebayashi, Akiko and Shun'ichiro Takagi. 1965. Gakurei mae no kodomo no kōon nōryoku ni kansuru ichikenkyū [A study on the articulatory ability of pre-schoolers]. *Onsei Gcngo Igaku* 6. 17–18.

Uwano, Zendo. 1999. Classification of Japanese accent systems. In Shigeki Kaji (ed.), *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics'*, 151–186. Tokyo: ILCAA.

Vallabha, Gautam K., James L. McClelland, Ferran Pons, Janet F. Werker and Shigeaki Amano. 2007. Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences* 104(33), 13273–13278.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Vance, Timothy J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Vance, Timothy J. this volume. Chapter 10: Rendaku. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Vihman, Marilyn May. 1996. *Phonological development: The origins of language in the child*. Oxford: Blackwell.

Warner, Natasha and Takayuki Arai. 2001. Japanese mora-timing: A review. *Phonetica* 58(1–2). 1–25.

Weber, Christiane, Anja Hahne, Manuela Friedrich and Angela D. Friederici. 2004. Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research* 18(2). 149–161.

Weismer, Gary and Mary Elbert. 1982. Temporal Characteristics of "functionally" misarticulated /ʃ/ in 4- to 6-years-old children. *Journal of Speech, Language and Hearing Research* 25(2). 275–287.

Werker, Janet F., Ferran Pons, Christiane Dietrich, Sachiyo Kajikawa, Laurel Fais and Shigeaki Amano. 2007. Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103(1). 147–162.

Werker, Janet F. and Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7(1). 49–63.

Yukari Hirata

18 L2 phonetics and phonology

1 Introduction

This chapter introduces research on phonetic and phonological aspects of Japanese as learned by non-native speakers of Japanese, and contextualizes it in the broader field of second language (L2) speech acquisition. Four sections introduce extant research in elements of Japanese speech sounds that are difficult for L2 learners to produce and perceive: single and geminate consonant contrasts (section 2), short and long vowel length contrasts (section 3), pitch accent (section 4), and stop voicing contrasts (section 5). Section 6 introduces extant theories of L2 speech acquisition, in which various factors predict degrees of success in L2 acquisition. Section 7 reviews studies exploring different types of training and multimodal-learning methods to enable learners to acquire difficult L2 speech sounds. The final section discusses areas of L2 Japanese that need investigation in the future.

L2 phonetics and phonology of Japanese is a highly interdisciplinary field in which researchers aim at understanding how non-native speakers learn to speak Japanese with native-like pronunciation in their life span. It is a complex question requiring a wide range of methods and approaches across different fields. As a start (sections 2–5), we need accurate descriptions of L2 Japanese learners' perception and production performance with a given native language (L1) background, and need to gain comparative understanding of different degrees of difficulty that learners face with different L1 backgrounds. This line of research often involves examining perception and production of native Japanese speakers (NJs) as a point of comparison. We also need to gain understanding of learners' developmental stages as they progress in their study of Japanese from beginning to advanced and near-native levels, and determine which elements of spoken Japanese are quickly, eventually, or never successfully learned. Even when we limit our investigations to adult learners, it is a challenge for L2 researchers to tackle questions such as how well age of learning, length of exposure, and L1 background can predict learners' ultimate attainments, and why L2 learning almost always involves large individual variations even if those factors are controlled. To investigate all of these questions, instructors of Japanese for L2 learners have taken anecdotal approaches with which they make valuable observations and state their intuitions, and L2 researchers have taken empirical and experimental approaches with which they confirm anecdotes with valid behavioral data.

Research in Japanese as an L2 has a long history in Japan, dating back to the 1940s, when Japanese was taught in various countries in Asia (Irie 1941). Kurono (1941), for example, noted problems of Japanese stop voicing distinction and obstruent length distinction for learners in Asia. Yamada (1963) conducted one of earliest

experiments with two Korean learners of Japanese, in which he examined the nature of their perceptual difficulty with vowel length and stop voicing distinctions. Since the 1960s, various problems have been noted for each L1 group (e.g., English: Kindaichi 1963; Takakura 1988; Taiwanese: Nozawa 1974; Cantonese: Hasegawa 1977; Chinese in general: Liu 1984; Indonesian: Sato 1986; Thai and Vietnamese: Yasuhara and Kasuya 1994; Korean: Matsuzaki 1999; Brazilian Portuguese: Sukegawa 1999). These descriptions, though they are often anecdotal, have become important steps and have given ideas for more tightly controlled experiments.

Moraic length contrasts, such as single and geminate obstruents and short and long vowels, as well as pitch accent, are the elements of L2 Japanese that are difficult for a large number of groups of learners to acquire, and which have been, by far, the most widely studied in and outside Japan (Kashima 2003) (see Otake, this volume, for full discussion of mora in Japanese). In addition, stop voicing contrasts have been studied extensively in Japan, and these are observed and tested to be some of the most challenging elements for learners of Japanese whose L1s are various Asian languages. Overviews on these well-studied elements are given in Ayusawa (2001), Sukegawa (2001), Oguma (2002), Kashima (2003), Toda (2003, 2009), and Hirata (2009). Sections 2–5 of this chapter introduce specific experimental findings on these elements in terms of their perception and production, and with regard to effects of phonetic contexts, syllable structures, word-internal positions, effects of speaking rate, effects of L1 backgrounds, and developmental stages.

Theories of L2 speech acquisition aim at providing possible mechanisms that are responsible for successful learning and development, as well as predicting what hinders specific phonetic and phonological learning. The L2 acquisition theories and hypotheses introduced in section 6 are proposed not only for Japanese, but they are constructed to account for general mechanisms for acquisition of different languages of L2s by different learner groups. They provide a helpful guide in accounting for converging and diverging patterns of L2 acquisition across languages. Finally, L2 researchers, as well as Japanese language instructors, have keen interest in knowing what kind and amount of input and practice would yield what degrees of success in learning (section 7). The topic of training is one of the richest areas of the L2 acquisition research because answers to this question have not only theoretical but also practical implications.

There are numerous other topics, however, that deserve attention but that have not been studied as extensively as those introduced in this chapter. Unfortunately due to the space limit, this chapter does not cover topics that have been studied but not yet developed. Those include acquisition of the overall moraic rhythm in phrases and sentences (Mizutani 1976; Kawakami 1984), two moras as a unit to be learned (Nakamichi 1980; Kanda and Uozumi 1995), sentential intonation, and segmental contrasts other than those mentioned above (see section 8 for future studies). Also not included are practical suggestions as to which phonetic and phonological elements should be taught explicitly in language curricula and how they should be

introduced in textbooks (Toki 1986), and social and cultural factors that might affect the degree of success in acquisition of Japanese as an L2 (Sukegawa 2001).

2 Acquisition of single and geminate consonant contrasts

L2 learners' acquisition of Japanese single and geminate consonant contrasts is one of the topics that are best studied in the field of L2 phonetics and phonology of Japanese. This section reviews empirical studies that examined this phonetic element separately for its perception (section 2.1) and production (section 2.2). See Kawahara (Ch. 1, this volume) and Kawagoe (this volume) for the phonetic and phonological nature of geminate obstruents in Japanese, and Ota (this volume) for the acquisition of consonant length contrast by native Japanese infants.

2.1 Perception

The major perceptual cue that native Japanese speakers (NJs) use to distinguish Japanese singleton and geminate obstruents is supposed to be their duration (Fujisaki, Nakamura, and Imoto 1975; Otsubo 1981; see Kawahara, Ch. 1, this volume, for details). However, as reviewed in section 2.1.1, non-native learners of Japanese, especially at their early stages of learning, may use different cues, e.g., those which are used to perceive obstruents in their L1. Their perception is also affected by factors such as lexical pitch accent, phonetic contexts, and speaking rate, which are reviewed in sections 2.1.1–2.1.3. As reviewed in section 2.1.4, however, their perception moves towards that of NJs as they advance their study of Japanese. Finally, section 2.1.5 reviews how learners of Japanese with different L1s differ in their perception.

2.1.1 Perceptual cues

Min (1987) investigated Korean learners' perception of Japanese singleton and geminate stops, and compared it with NJs' perception. Learners who had been studying Japanese for about 30 hours in high school in Korea did not identify stimuli based on duration of stop closure as NJs did, but their perception was largely affected by phonetic characteristics of the stops. Min describes that Japanese geminates are produced with a short aspiration and tensed larynx, which are perceived by Korean learners as Korean tense unaspirated stops [p' t' k']. Min suggests that it is important to explicitly teach not to substitute Japanese geminates with these

Korean tense unaspirated stops but rather to pay attention to stop closure duration. Horigome (1999) examined five Korean learners' perception of singleton and geminate stops, and also found that their perception was affected by quality (instead of duration) of the stops. In Hung (2012), 312 Taiwanese learners identified single and geminate stops based on the stop closure duration, but their identification of geminates increased linearly with increases in duration. This contrasted with NJs' identification that was more categorical, showing less ambiguity and a sharper perceptual boundary between singletons and geminates.

Lexical pitch accent is known to affect the difficulty in the perception of stop length contrasts. Minagawa and Kiritani (1996) found that, for HL (high and low) words (e.g., [ha'to] 'pigeons' and [ta'tte] 'stand up'), Korean and Chinese learners misperceived singleton stops as being geminates significantly more than geminate stops as being singletons. However, for LH (low and high) words (e.g., [nata] 'wide blade knife' and [satto] 'quickly'), the rate of misperception did not differ between the words with singletons and those with geminates. Acoustic measurement revealed that the second vowels of the disyllables were consistently shorter in the HL than LH words, while the closure duration did not differ between the two accent types. Minagawa and Kiritani discussed that these learners may be using the durational ratio of stop closure to the following vowel as a perceptual cue. This is an interesting difference between these learners and NJs because NJs are known to use the ratio of stop closure to the preceding vowel (C/V1) (Watanabe and Hirato 1985; Hirata 1990a), but not to the following vowel (Hirato and Watanabe 1987), as a perceptual cue when disyllables are spoken in isolation.

2.1.2 Effects of phonetic contexts

Hardison and Motohashi (2010) examined perception of singleton and geminate consonants by native English learners of Japanese, including 28 beginning, 42 low–intermediate, and 15 advanced learners who were in the first, the third, and the seventh semesters of Japanese language study. They found that identification accuracy was significantly lower for words including geminate fricatives [ss] than stops [kk tt], e.g., [sassɯ] versus (vs.) [sakkɯ], and lower for words in which the vowel following geminate [ss] was [ɯ] than [a], e.g., [sassɯ] vs. [sassa]. This pattern of results was stronger when these words were presented in a carrier sentence than in isolation. They explained that this difference in identification accuracy is attributed to the sonority differences of the examined segments, i.e., low vowel [a] > (greater than) high vowel [ɯ] > fricative [s] > stops [t k] (Wright 2004). It is easier to accurately identify the consonant length distinctions when the sonority differences are greater between the preceding vowel and the target consonant contrast, i.e., [a-k] > [a-s], and between the target consonant contrast and the following vowel, i.e., [s-a] >

[s-ɯ]. In these example pairs, the larger the sonority differences, the larger the perceptual distance and therefore the more accurate the perception of the consonant length. The result that stop length contrasts are easier to acquire than fricative length contrasts replicated earlier findings by Toda (1998a, 2003) (see section 2.1.4). It is interesting to point out that NJs do not seem to be affected by the sonority factor (Fujisaki, Nakamura, and Imoto 1975), but that non-native learners of Japanese are susceptible to it.

2.1.3 Effects of speaking rate

It is known that NJs adjust their perception according to speaking rate. Hirata (1990b) found that NJs identified length of a Japanese stop in [ita]-[itta] 'stayed'-'went' based on the durational relationship between the stop closure and the preceding vowel when the words were presented in isolation. However, their perception was adjusted according to speaking rate of the materials following the target word in the sentence [i(t)ta koto to i:maʃita]. The faster the portion following the stop closure, the more responses were given to the geminate, and the singleton or geminate stops were perceived categorically. This ability to categorically perceive length contrasts based on speaking rate of sentences was not found for native English beginning learners of Japanese (Hirata 1990b). When hearing isolated words, native English learners of Japanese were also found not to adjust their perception according to duration of a preceding vowel as NJs do (Toda 1998a).

Studies have accumulated to show that it is a challenge for learners of Japanese to constantly take speaking rate into account in perceiving Japanese length contrasts (e.g., Sonu et al. 2011a; Sonu et al. 2012; Sonu et al. 2013). Sonu et al. (2011a) found that, for native Korean speakers who had studied Japanese for 170 hours, there was a perceptual bias toward hearing single obstruents as geminates particularly when words spoken at a slow rate were presented in isolation. This challenge of coping with speaking rate variation is discussed more in section 7.1.2.

2.1.4 Developmental stages

Enomoto (1992) examined perception of length contrasts by learners at three levels: beginning (6 weeks of Japanese study; n=6), intermediate (about a year; n=6), and near-native ("substantial experience of using Japanese in Japan", Enomoto 1992: 30) (n=2). Stimuli were continua varying in consonant duration in, e.g., [i(k)ken] and [ni(ʃ)ʃi]. Results of these learners were compared to those of NJs. The NJs' perception was categorical, i.e., perceiving continua with shorter and longer duration clearly either as singletons or geminates. The beginning learners did not show this

sharp categorical perception, and even the end-point stimuli of [iken] and [ikken] (where the NJs showed 100% identification of either the "singleton" or the "geminate" response) were not perceived at a 100% rate. However, there was a tendency for the intermediate and near-native learners to reach the NJs' categorical perception pattern, indicating that learners do make progress in their perception of these contrasts as their language experience increases.

Toda (1998a) compared beginning Japanese learners (college students) and advanced Japanese learners (diplomats, no other information provided) with NJs in their perception of edited disyllable continua, such as [kate]-[katte] and [iso]-[isso], varying in duration of the word-medial consonants. NJs identified these stimuli categorically and their categorical boundaries did not differ whether the stimuli were presented in the order of short to long (ascending) or long to short (descending). However, the beginning Japanese learners' categorical boundaries differed significantly between the two sequence conditions. In the descending sequence they switched their identification from "geminate" to "singleton" when the closure duration was still in the "geminate" category for the NJs, and their mean categorical boundaries were significantly greater than the NJs. For the advanced Japanese learners, this discrepancy disappeared on *stop* contrasts, showing categorical boundaries similar to those of the NJs. However, their categorical boundaries were significantly different from the NJs on the *fricative* and *nasal* pairs. Thus, the advanced learners did make a progress in their perception of stop length, but not of the fricative and nasal length contrasts.

2.1.5 Effects of L1

Minagawa (1996) and Minagawa and Kiritani (1996) compared perception of Japanese single and geminate stops by learners with five L1 backgrounds: Korean, Thai, Chinese, English, and Spanish. Three interesting results were found. First, overall perception accuracy was significantly higher for the English and Spanish groups than the others. Second, overall error patterns showed that all language groups except for the Spanish group had dominant misperception of single stops as having geminates compared to the other way around. Third, the Korean and Chinese groups showed similar error patterns associated with lexical pitch accent as mentioned in section 2.1.1: Singletons were misperceived more as geminates in HL, but LH yielded the same amount of misperception for singletons and geminates. In contrast, the Thai group's misperception was not affected by lexical pitch accent, and had a tendency to perceive more singletons as geminates regardless of the pitch accent differences. Minagawa (1996) discusses that Korean, Chinese, and Thai have phonemic contrasts in stop aspiration, and they might use some acoustic features of Japanese stops other than closure duration as a perceptual cue to the length contrasts.

2.2 Production

Production of Japanese singleton and geminate obstruents is also a challenge for L2 Japanese learners. Their production is affected by the articulatory and timing controls that are used in their L1, which is reviewed in sections 2.2.1 and 2.2.3. Although some studies show that their production improves as they advance their Japanese study, there are other studies showing that even advanced learners have difficulty producing these contrasts accurately (section 2.2.2).

2.2.1 Articulatory and timing controls

Toda (1997) examined initial strategies that native English learners of Japanese use for the production of geminate stops. In English, the only phonetic context in which a geminate consonant appears is at a word boundary, e.g., *white tie*, and this affects the learners' initial strategy for producing Japanese geminates as CVC#CV (i.e., a closed heavy syllable followed by an open light syllable). Their production of Japanese [rikka] was manifested in two strategies: (1) showing a small release of a stop [k] for the first syllable [rik], and (2) lengthening the first vowel [ri:k] in an attempt to produce a syllable of long duration.

Han (1992) compared production of singleton and geminate stops by NJs and native English advanced learners of Japanese. NJs' productions showed a mean ratio of geminates and singletons as 2.8:1.0, whereas the learners showed approximately 2.0:1.0, pointing out the challenge of this distinction even at an advanced level.

2.2.2 Developmental stages

Masuda and Hayes-Harb (2005) conducted acoustic analysis of single and geminate stops produced by native English speakers. NJs' ratios of word-medial obstruents (stop [t] or fricative [s]) to the preceding vowels (C/V1) were 1.91 for singletons and 3.65 for geminates. While native English speakers with no experience with Japanese language study showed ratios for singletons and geminates that were not as clearly separated (1.23 vs. 1.58), intermediate learners with one year of Japanese language study reached values (2.07 vs. 2.98) close to those of NJs. Masuda and Hayes-Harb (2007) found similar results for Korean learners of Japanese: Intermediate-level learners (with 6–24 months of Japanese study) were more accurate and showed scores closer to NJs than beginning learners (with less than 6 months of Japanese study).

2.2.3 Effects of L1

Masuda (2009) examined Japanese single and geminate stops in disyllables produced by native English and native Korean learners of Japanese at beginning and intermediate levels, who had been studying Japanese for 6–12 months and about two years, respectively. The native English learners' differences in C/V1 ratios between singleton and geminate stops were not large compared to those of NJs, and they did not improve in this distinction from the beginning to the intermediate levels. In contrast, the Korean learners showed C/V1 ratios that were larger than those of NJs for both singletons and geminates, and there was no clear improvement from the beginning to the intermediate level. Masuda concluded that learners of different L1s use different production strategies.

Korean learners of Japanese are known to incorrectly produce Japanese intervocalic single voiceless obstruents as geminates. Min (2007) gave detailed explanations about how they use Korean tense obstruents as a substitute for the intervocalic single voiceless obstruents, showing an effect of their L1 phonological system. Min provided persuasive production data for Korean two-syllable sequences, showing that they produce the first syllable as a closed syllable with the first vowel shortened, and as a consequence, the closure duration increases to the extent that NJs would hear a geminate.

3 Acquisition of short and long vowel length contrasts

Japanese short and long vowel length contrast is another phonetic element that is known to be difficult for non-native speakers to acquire (see Ota, this volume, for the acquisition of vowel length contrast in L1 phonology). As in the previous section, empirical studies regarding this phonetic element are reviewed separately for perception (section 3.1) and production (section 3.2). Furthermore, researchers have been interested in comparing learners' acquisition of vowel length as opposed to consonant length contrasts, given that these two phonetic elements have the same underlying mechanism for NJs (Fujisaki, Nakamura, and Imoto 1975). This is discussed in section 3.3.

3.1 Perception

Factors known to affect difficulty in perceiving vowel length contrasts are word-internal positions, lexical pitch accents, segmental contexts, and speaking rates; these are reviewed in sections 3.1.1, 3.1.2, and 3.1.3. How learners make progress

with their perception and how their L1s affect their perception are discussed in sections 3.1.4 and 3.1.5.

3.1.1 Effects of word-internal positions and pitch accents

One of the earliest studies in investigating the difficulty for learners of Japanese to distinguish Japanese short and long vowels is Yamada (1963). Yamada conducted an experiment with two Korean learners of Japanese, and found that the distinction of short and long vowels was not hard when the contrast appears in word-initial syllables, but harder in word-medial syllables, and even harder in word-final syllables. The results of this small-scale experiment were confirmed by larger-scale studies more recently. Minagawa, Maekawa, and Kiritani (2002) examined effects of word-internal position and pitch height on learners' perception of Japanese short and long vowels. This study involved a large pool of Japanese learners, 30 native English and 30 native Korean speakers whose length of Japanese study was balanced in each group: 40% had 3–12 months, another 40% had 1–2 years, and the remaining 20% had 2–3 years of Japanese language study. Participants were presented with both nonsense and real disyllabic words and identified whether each syllable included a long vowel. For both of the learner groups, more errors were made when vowels were in word-final than word-initial syllables. For words including a long vowel in word-final position, pitch accent patterns of the words had a large effect: more errors were made when the long vowel was in the HLL pitch pattern than in the LHH or LHL patterns. For words including only short vowels, the final vowel in the LH pitch pattern was misidentified as being a long vowel. Minagawa et al. explained that higher fundamental frequencies in the high tone psychoacoustically tend to be heard as longer sounds, and that learners of Japanese were biased by this psychoacoustic tendency.

In addition, Minagawa (1997) found that the results described above showing effects of word-internal positions and pitch accents were also found in learners whose L1s are Thai, Chinese, and Spanish. Regardless of these L1 backgrounds, long vowels in the HH pitch pattern tended to be identified correctly and those in the LL pattern tended to be misidentified as short vowels. On the other hand, short vowels with an H tone are misidentified as long vowels and those with a L tone are correctly identified as short. For all listener groups, identification in word-final position yielded the highest errors. Najoan et al. (2012) obtained similar results for native Indonesian learners of Japanese with regard to long vowels in word-final position with an LL pattern showing the highest errors.

3.1.2 Effects of segmental contexts

One factor that has not been studied thoroughly is the effect of segmental variations on non-native listeners' perceptual accuracy for Japanese vowel length distinctions.

A preliminary study by Nakagawa and Futamura (2000) shows that in the perception of disyllables (with 14 different native languages) short vowels preceded by [j], e.g., [jokɯː], tended to be misperceived as long more than those preceded by other consonants, e.g., [tokɯː]. Thus, only small perceptual errors were found for long vowels such as [joːkɯ] or [kɯjoː]. There seem to be complex interactions among the factors of the consonants (e.g., [j] vs. [Cj] vs. others) that precede contrasting vowels, word-internal positions (first vs. second vowels in disyllables), and pitch accents. Further research is necessary to pinpoint how each of these factors interacts with each other.

3.1.3 Effects of speaking rates

One of non-native speakers' difficulties involved in the distinction of Japanese short and long vowels is that the phonemic categories of "short" and "long" depend on the rate at which utterances are spoken. It is known that long vowels are 2.4–3.2 times longer than short vowels in NJs' production, but when speech varies from slowest to fastest speaking rates, there is significant overlap in absolute duration between long vowels spoken quickly and short vowels spoken slowly (Hirata 2004a). Hirata (2004a) found that, despite this overlap in speech of varied rates, the durational ratio of a target vowel to a disyllable is stable acoustic information to reliably classify the two phonemic categories of "short" and "long". As a result, NJs have no problem adjusting to speaking rates and perceive vowel length accurately when sufficient surrounding contexts are present (Hirata and Lambacher 2004).

Although there is thus clear durational information within an utterance and NJs efficiently use that information to identify vowel length across different speaking rates, non-native speakers have difficulty in doing so. Tsurutani (2003) examined how accurately native English learners of Japanese perceived both vowel and stop length distinctions in Japanese disyllables spoken at slow and fast speaking rates. Results showed that both beginning and advanced learners misperceived slow-rate CVCV stimuli as having a geminate stop, and misperceived fast-rate CVVCV or CVCCV stimuli as having a short vowel or a singleton stop. These results indicate that learners were affected by absolute duration of the stops and unable to adjust their perception according to speaking rate variations. However, the advanced level learners made few errors in identifying long vowels and geminate stops in fast speech.

3.1.4 Developmental stages

Toda (1998a) compared beginning Japanese learners (college students) and advanced Japanese learners (diplomats, no other information provided) with NJs in their perception of edited disyllable continua varying in duration of the second vowels, e.g.,

[kate]-[kate:]. NJs' categorical boundaries did not differ whether stimuli sequences were presented in the order from short to long duration (ascending) or long to short duration (descending). However, the beginning Japanese learners' categorical boundaries differed significantly between the two sequence conditions: in the descending sequence they switched their identification from "long" to "short" when the vowel duration was still in the NJs' "long" vowel category. However, this discrepancy disappeared for advanced Japanese learners, showing categorical boundaries similar to those of NJs. These results are similar to those reported earlier on single and geminate stops (section 2.1.4).

Oguma (2000) examined how the perception of Japanese short and long vowels in two- to four-mora words develops as learners progress in their language proficiency. In this study, native English learners of Japanese were divided into three groups of beginning, intermediate, and advanced groups, based on their proficiency roughly corresponding to levels 4, 3, and 2 of the world-wide Japanese Language Proficiency Test administered by the Japanese government. The perception scores were significantly different between the intermediate and the advanced levels, suggesting that the improvement occurs around the transition between these stages, but not earlier. For all three groups, there were more misperceptions of long vowels as short than short vowels as long. The misperceptions of short vowels decreased at the intermediate level, but those of long vowels did not decrease until at the advanced level. Consistent with Minagawa, Maekawa, and Kiritani (2002), long vowels in the LL pitch pattern were more difficult to accurately perceive than those in HL, HH, or LH, but perception improved at the advanced level.

3.1.5 Effects of L1

Minagawa (1997) compared the perceptual patterns of Japanese short and long vowels among native speakers of Thai, Chinese, Spanish, Korean, and English. While there was commonality in their perception patterns across the five language groups (with regard to the effects of position and pitch accent as described in section 3.1.1), differences were also found. For example, the Thai group made fewer errors than the Chinese group because Thai, but not Chinese, has a vowel length distinction. The Spanish group made more errors than the English group, which Minagawa pointed out as needing further investigation and explanation.

Kurihara (2007) examined effects of L1 backgrounds with Japanese language learners who were native speakers of Finnish, Chinese, and Korean. When presented with continua varying in duration of the first and the second vowels of a Japanese disyllable, the Finnish learners perceived these stimuli in a categorical manner similar to that of NJs. In contrast, the Chinese and Korean learners of Japanese showed identification patterns that were less categorical. This difference between the identification patterns by the Finnish and the other language groups was more exaggerated

when the vowels were in the second syllable (i.e., word-final position) than in the first syllable. Finnish has vowel length distinctions that are similar to those of Japanese, and this may explain why the Finnish group had an advantage in perceiving Japanese vowels.

A similar finding was obtained by Tsukada (2011a), in which Japanese vowel length contrasts were identified by native Arabic and Persian speakers who had no prior experience with Japanese. Arabic has vowel length distinctions but modern Persian does not, and the Arabic group showed identification accuracy similar to that of NJs, significantly better than the Persian group. It is noteworthy that this result was not observed when they were tested with an AXB discrimination task, in which participants were presented with a three stimulus sequence (A, X, and B) and judged whether the second stimulus (X) was the same as the first (A) or the third (B). This task was known to require different demands where L1 linguistic categories in the long-term memory may not necessarily be referred to.

3.2 Production

Below are two subsections on non-native learners' production of short and long vowels. Similar to the case of perception (section 3.1.4), learners improve in their production as they advance their study of Japanese, but there always seems to be individual variation in the degree of improvement (section 3.2.1). Learning to produce these vowels accurately at varied speaking rates is a challenge for learners (section 3.2.2), which also echoes the case of perception (section 3.1.3).

3.2.1 Developmental stages

Oguma (2001) examined learners' production of two- to four-mora Japanese words that included short and long vowels in various positions. As in Oguma (2000) (section 3.1.4), native English learners of Japanese were divided into three groups of beginning, intermediate, and advanced groups, based on their proficiency corresponding to levels 4, 3, and 2 of the Japanese Language Proficiency Test. For words produced in isolation, production accuracy evaluated by NJ judges did not differ among the three levels of learners, but for words produced in sentences, accuracy was significantly higher for the intermediate and advanced level learners than beginning learners. Oguma (2001) attributed this difference to the attention to materials. Note that other factors may have well been involved. In listening to learners' test sentences, e.g., [ato de joi kasa o joːi ʃite kudasai] 'Please have good umbrellas ready later,' NJ judges may have identified the words [joi] 'good' and [joːi] 'ready' based on the semantic contents, and not based on the duration of the contrasting vowels *per se*. Thus, the higher-level learners' overall fluency, including rhythm, speaking rate,

sentential intonation, as well as appropriate production of all individual segments may have helped for overall sentence intelligibility, which resulted in higher scores on the target vowel length. This is an interesting result that should be followed up in the future.

Ueyama (2012) conducted a production experiment with four intermediate (2.5–6 years of Japanese study with no experience living in Japan) and three advanced learners (3–9 years of Japanese study with 4–11 years of experience living in Japan). Test words were disyllables in which the first vowels contrast in length, e.g., [birɯ]-[bi:rɯ], produced in a semantically-neutral carrier sentence [soʃite ___ to i:maʃita]. Three NJs' ratios of long to short vowels were 2.0–2.3, and one advanced and one intermediate learners' ratios successfully fell within this range. The ratios of the other two advanced and two intermediate learners were above the native speaker range (i.e., overshooting), and one intermediate learner showed undershooting. Ueyama concluded that learners varied in their abilities to produce native-like short and long vowels, and their abilities did not correlate with the number of years of Japanese study or experience of living in Japan.

3.2.2 Effects of speaking rate

Yi (2003) compared the durations of Korean learners' and NJs' production of Japanese short and long vowels spoken at different speaking rates. Yi found that the Korean learners produced words with the same number of syllables with almost the same duration, whereas the NJs produced them according to the number of moras in the words. Another difference between the two groups was that the NJs had stable durational ratios of the first vowel to the second vowel, whereas the Korean learners did not, indicating the learners' unstable productions when they speak at slow and fast rates.

Jia, Mori, and Kasuya (2005) conducted acoustic analyses of learners' production of vowel and consonant length contrasts, [gosekɯ]-[gose:kɯ]-[gosekkɯ], spoken in a carrier sentence at four speaking rates, and compared the results with those of NJs. They measured the added duration for an additional mora from the long vowel in [gose:kɯ] or from the geminate stop in [gosekkɯ], each compared with [gosekɯ]. For the NJs, this added duration changed linearly as speaking rate changed, with little variability among the speakers, suggesting that they have a stable durational control for moras. In contrast, native Chinese learners of Japanese showed much higher variability in realizing this durational pattern, and only one of the five learners showed durational control similar to that of the NJs. Similar to Yi (2003), Jia, Mori, and Kasuya (2005) pointed out the learners' instability in the production of moraic rhythm when they speak with varied speaking rates.

3.3 Comparisons of consonant vs. vowel length contrasts

It has been shown experimentally that NJs identify both consonant and vowel length contrasts categorically in a similar fashion, and that the perception of these two types of contrasts is driven by the same mechanism involving the processing of speech duration (Fujisaki, Nakamura, and Imoto 1975; Enomoto 1992). How about non-native speakers?

Masuko and Kiritani (1990) conducted a perception study with Chinese, Korean, Thai, and Indonesian learners of Japanese. These learner groups (n = 6–10 for each group) had, on average, 2 years (Chinese), 1 year (Korean), and 4 months (Thai and Indonesian) of experience studying Japanese. Stimuli were vowel length and consonant (stop, fricative, and nasal) length pairs, e.g., [kodai]-[ko:dai] and [itʃo]-[ittʃo]. Perception accuracy was found to be higher for vowel length than consonant length pairs for all groups, but this gap seemed to differ across the learner groups: the differences in percentage points were about 10–15% for the Chinese, Thai, and Indonesian groups, but 4% for the Korean group. The Korean group's performance was lowest for both the vowel and consonant pairs, while the other three groups did much better on the vowel pairs. Note that the vowel length contrasts in this study were all in word-initial syllables, which is the position within words that is easiest to perceive (section 3.1.1). Thus, this result *per se* does not show that vowel length contrasts are easier to perceive than consonant length contrasts.

How does the perception of vowel length contrasts in the most difficult position (i.e., word-final position; section 3.1.1) compare with that of the consonant length contrasts? In Toda (1998a), advanced learners' categorical boundaries for vowel length contrasts in this most difficult position were very similar to that of NJs, but the learners' boundaries for fricative and nasal length contrasts were significantly smaller than those of NJs. Among the consonant contrasts, the learners seemed to be best in perceiving stop length, while lagging behind with fricative and nasal length contrasts. Taken together, vowel length seems to be easier to perceive than consonant length.

Hirata and McNally (2010) examined *production* of vowel length vs. consonant length contrasts in [rika]-[rika:] and [kako]-[kakko] by seven native English learners of Japanese at an intermediate level (with 2 years of Japanese study in the U.S.) before and after their first-time four-month stay in Japan. Their production was analyzed acoustically and was compared to NJs' in terms of duration of the contrasting consonant and vowel, the ratio of those consonant or vowel to the disyllabic word, and the ratio of the two- and three-mora words. In all of these measures, the learners improved on the vowel length, but not on the consonant length production after this four-month stay in Japan. It is consistent with Toda (1998a) that the learners improved on the vowel pairs even in the most difficult word-final position, while they did not on the consonant pairs.

While the three studies above suggest that vowel length contrasts are easier to learn than consonant length contrasts in both perception and production, there are other studies that do not show this. Enomoto (1992) (as mentioned earlier) conducted small-scale categorical perception experiments on stop, fricative, nasal, and vowel length pairs by beginning (6 weeks of Japanese study), intermediate (1 year or less), and near-native learners ("substantial experience of using Japanese in Japan", Enomoto 1992: 30) who are native speakers of English. For all of the length contrast pairs, the learners' categorical perception patterns generally became closer to those of NJs as they moved from the beginning to the advanced levels.

At the very initial stage before starting Japanese study, learners' perception accuracy does not seem to differ between vowel and consonant (obstruent) length contrasts. Hirata (2004b) examined native English speakers' perception of Japanese length contrasts by asking them to count the number of moras in one- to six-mora words presented in isolation and in sentences. Having received the instruction on Japanese moras, the native English participants with no experience with Japanese were initially unable to detect long vowels and geminate consonants, and counted syllables. For example, words such as [tsɯbo] 'a pot' and [otosɯ] 'to drop' were correctly identified as containing two and three moras, respectively, but disyllabic words such as [boko:] 'alma mater' (3 moras), [ʃɯsse] 'career advancement' (3 moras), and [ke:se:] 'formation' (4 moras) tended to be miscounted as having two moras. The number of errors made for the long vowels and the geminate consonants did not differ at this initial stage, and even after ten sessions of intensive perceptual training, the amount of improvement also did not differ between words with long vowels and geminate consonants.

Hirata and Ueyama (2009) compared the native English speakers' results above with those of native Italian speakers. English has neither vowel nor consonant length contrasts as Japanese does, while Italian has consonant, but not vowel length contrasts. With no prior experience with Japanese, the Italian speakers' results did not differ from the above English speakers' on the long vowel words, but the Italian speakers showed significantly higher accuracy than the English speakers on the geminate words. However, the Italian speakers' advantage appeared only in sentences as opposed to isolated words. The results are noteworthy in two respects: First, this is a case in which an L1 effect superseded the general tendency of vowel pairs being perceived more easily than consonant pairs, and second, Italian speakers' unique advantage appeared in the sentence, but not in the isolated-word, contexts.

In summary, vowel length contrasts tend to be easier to learn than consonant length contrasts in terms of both perception and production, and the former is successfully learned by the time learners reach an advanced level, whereas the latter continues to pose difficulty even for advanced learners. However, this tendency may depend on, or interact with, the kinds of tasks given and the learners' L1 backgrounds.

4 Acquisition of lexical pitch accent

Lexical pitch accent is known to be another element, along with consonant length and vowel length contrasts, to cause difficulty for non-native learners' acquisition of Japanese (see Kawahara, Ch. 11, this volume, for the phonetics and phonology of word accent; Ota, this volume, for the acquisition of pitch accent by native Japanese speakers). The pitch accent is acoustically manifested in fundamental frequency (F0). As in the previous sections, empirical studies in perception and production of lexical pitch accent are reviewed separately in sections 4.1 and 4.2, respectively.

4.1 Perception

Questions addressed in this subsection include which accent patterns of Japanese words are difficult to perceive, how the learners' L1 systems (e.g., lexical stress in English and lexical tones in Chinese) affect their difficulty in perceiving Japanese pitch accent, and how they develop their perceptual ability over time (section 4.1.1). Other factors examined are syllable structures (section 4.1.2) and testing methods of discrimination versus identification (section 4.1.3).

4.1.1 Pitch accent types, L1 effects, and learning

Toda (2001) examined perception of lexical pitch accent in isolated four-mora words by intermediate and advanced learners of Japanese whose native languages varied across 17 different languages. Their scores were found to be initially highest for H'LLL and LH'LL, followed by LHHH, and lowest for LHH'L. After some instruction and training, the learners generally improved on the perception of H'LLL and LHHH, but perception of LHH'L stayed most difficult.

Ayusawa (2003) summarizes a decade of research she and her colleagues conducted in 1994–2003 regarding perception of Japanese lexical pitch accent by learners who vary in their language proficiency levels and their native languages. In Nishinuma, Arai, and Ayusawa (1996), 54 native English speakers who had studied Japanese for 2 years perceived lexical pitch accent in 3-5 mora words and phrases in carrier sentences. The pitch pattern LHHHH with no accent was perceived most correctly, and LHH'LL and LHHH'L were most poorly perceived, and LH'LLL and H'LLLL in between. Longitudinal studies were also conducted with learners of more than 15 different native languages, in which the learners improved their overall perception accuracy over varying amounts of time. Ayusawa (2003) summarizes that the degree of perceptual accuracy for different pitch accent patterns depends on learners' native languages, and that the learners show higher accuracy in those

pitch accent patterns that are similar to the prosodic patterns of their L1. She also points out that perceptual accuracy varies across different individuals, regardless of their history of Japanese language study.

Hirano-Cook (2011) conducted a large-scale perception experiment with native English learners of Japanese across five levels (33, 31, 24, 26, and five participants at each of the first through the fifth year levels, respectively, of Japanese study at a university in the United States). The learners were asked to identify one of the four types of pitch accent patterns of four-mora words in isolation, and they perceived LHHH best, H'LLL least well, and LH'LL and LHH'L in between. This pattern of difficulty did not change much across the learners from the first-year to the fifth-year levels, although their accuracy increased as they gained more experience of studying Japanese. Hirano-Cook explains that the difficulty of H'LLL may relate to the characteristic of fundamental frequencies (F0), having an initial rise and a delayed peak.

In Shport (2011), 21 native English speakers were tested, 18 of whom had never learned Japanese before and three of whom had 1–6 years of Japanese study. The learners' perceptual ability did not differ among H'L+L, LH'+L, and LH+H word types presented in sentences at the first test, but half of the participants who initially scored low made improvement after one-hour training for H'L+L and LH'+L (but not LH+H).

Taken together, these studies do not point to clear common conclusions about the levels of difficulty in perceiving pitch accent types, partly because they tested different groups of Japanese language learners at different levels with different native language backgrounds. Stimulus factors such as whether words were presented in isolation or in carrier sentences must also play a role. Another factor is syllable structures, which is discussed below.

4.1.2 Syllable structures

In Hirano-Cook (2011), learners generally had more difficulty identifying pitch accent patterns of words that included heavy syllables than light syllables. Similar results were found in Toda (2001), but she also showed that instructions and training help learners to improve their perception of heavy syllables. Interestingly, the opposite pattern of results was found for Korean learners. Sukegawa and Sato (1994) conducted a discrimination test with advanced Korean learners with 4–10 years of Japanese study, and their discrimination of pitch accents was better in CVV ([ne:] and CVN ([nen]) than in CVCV ([nene]). There are a number of differences in the three studies above, e.g., discrimination vs. identification tests, number of moras in test words, edited vs. naturally spoken stimuli, and learners' native languages. This investigation in effects of syllable structures on learners' perception would be an interesting line of work in the future.

4.1.3 Discrimination vs. identification

With regard to learners' general difficulty in perceiving Japanese lexical pitch accent, one may ask whether the difficulty is caused by low-level, auditory, and psychophysical inabilities, or lack of higher-level, cognitive, and linguistic understanding of the accent system that operates categorically. This question can partially be addressed by studies that utilize non-speech stimuli that vary in F0 contours. Sakamoto (2010) used stimuli that sounded like buzzing noises, whose F0 contours resemble those of naturally spoken Japanese lexical pitch patterns of two-mora words followed by a particle, e.g., [mene mo]. Participants were "inexperienced" learners of Japanese who had an average of 2 years of Japanese study and 0–8 weeks of stay in Japan, and "experienced" learners who had an average of 3.7 years of Japanese study with 1 year of stay in Japan. The task was an ABX discrimination task in which three stimuli were presented in sequence for each trial, and listeners were asked to determine if the third stimulus was the same type of sound as the first or the second stimulus. Both groups did not have problems in discriminating the three types of non-speech F0 contours resembling H'L+L, LH'+L, and LH+H of natural speech. This result shows that learners' difficulty was not due to low-level, auditory, or psychophysical inabilities. However, a different result emerged when naturally spoken words of the three pitch patterns (H'L+L, LH'+L, and LH+H) were presented in a carrier sentence and participants were asked to identify them, i.e., asked to choose one of the three categories for each stimulus presented at a time. This task required them to form three abstract linguistic categories. The inexperienced group's score was significantly lower than NJs' score, but the experienced group's score did not significantly differ from that of the NJs. This result indicates that the experienced, but not the inexperienced, learners were able to form abstract linguistic categories.

A similar result was obtained by Hirano-Cook (2011) with learners who were at the 2nd- and 3rd-year level of Japanese study. She conducted an AX discrimination test, i.e., giving two stimuli and asking whether the second stimulus was the same or different from the first. Forty four-mora words with all light syllables were presented in pairs, and were judged whether two stimuli had the same or different accentual patterns. The learners were divided into two groups, "top" and "bottom", according to their identification test scores described earlier in section 4.1.1. The listeners in the bottom half group who were unable to identify the accent patterns of the four-mora words, were found to be able to score on the discrimination test as high as the listeners in the top half group.

The two studies above show that learners' inability to perceive the lexical pitch patterns accurately is not due to their inability to auditorily detect signal (F0) differences between two words, but due to their inability to categorize them according to the Japanese linguistic system.

4.2 Production

Studies investigating learners' production of Japanese lexical pitch accent are reviewed below. Factors examined are pitch accent types (section 4.2.1), developmental stages (section 4.2.2), syllable structures (section 4.2.3), and L1 tonal features (section 4.2.4). Investigations in these areas can be expanded further by examining a variety of subjects with different L1s, stimuli, and testing methods. It is hoped that the studies reviewed below will spark interest in further investigations.

4.2.1 Pitch accent types

In Sakamoto (2010), NJs and two groups of native English learners of Japanese, experienced and inexperienced (section 4.1.3), produced nonsense disyllables followed by a particle *mo* (e.g., [mene mo] and [nime mo]) in the three pitch accent types, H'L+L (A1), LH'+L (A2), and LH+H (A0). NJs then identified those recorded nonsense disyllables in terms of the intended pitch accent types. Both the experienced and inexperienced learners' scores were highest for A0, lowest for A2, and A1 in between. The experienced group's production scores were notably higher than the inexperienced group's, but still lower than the NJs'.

For both groups, the highest misidentification by NJs of the above production occurred with A1 or A2 perceived as A0, and the second highest misidentification was of A2 perceived as A1. Analyses of the learners' fundamental frequency contours revealed that the first type of misidentification occurred because the learners' contours were more flat than the A1 or A2 produced by NJs. The second type of misidentification occurred due to the learners' F0 peak coming earlier than that of the A2 produced by NJs.

4.2.2 Developmental stages

Sakamoto (2010) above further analyzed and compared production and perception results (section 4.1.3), and found that, while experienced learners do progress in their abilities both to produce and to perceive pitch accents, production seems to lag behind perception. The experienced learners with an average of 3.7 years of Japanese study with 1 year of stay in Japan showed perception of the three types of pitch accent similar to that of NJs, but their production was still significantly lower than that of native speakers.

4.2.3 Syllable structures

Sukegawa (1999) conducted a production experiment with two Brazilian learners of Japanese. The produced words were judged by NJs in terms of their pitch accent

patterns. It was found that the Brazilian learners' pitch patterns of CVNCV (e.g., [genki]) and CVVCV (e.g., [do:ro]) were heard by NJs as HHL instead of the correct HLL. The author interpreted this result as the learners assigning pitch accent based on syllables, not moras.

4.2.4 Effects of L1 tonal features

In Nozawa and Shigematsu (1998), Cantonese learners of Japanese tended to produce Japanese words with flat patterns, with smaller dynamic changes in F0 (i.e., less rising and falling contours of pitch accent). They attributed these patterns as being influenced by Cantonese tones, explaining that the younger generation of Cantonese speakers as in their experiment have a tendency to use more flat tonal patterns than older generations.

5 Acquisition of stop voicing contrasts

It has been more than half a century since Kurono (1941) noted the problem of native Thai, Chinese, and Korean learners of Japanese in accurately producing Japanese obstruent voicing contrasts. Since then, abundant empirical research has been conducted on L2 Japanese learners' difficulty for this phonetic element, but interestingly, most publications are in Japanese. Most studied are stop voicing contrasts perceived and produced by L2 Japanese learners whose L1s are Asian languages. In this section these studies are reviewed by learner groups: Korean learners of Japanese (section 5.1), Beijing Mandarin and Shanghainese learners (section 5.2), and learners of other L1 backgrounds (section 5.3).

The major acoustic correlate and perceptual cue to stop voicing distinction is voicing onset time (VOT), which is an interval between the onset of a burst release of lips and the onset of vocal fold vibration. Positive VOT values mean that the burst release is followed by vocal fold vibration, while negative VOT values mean that vocal fold vibration precedes the burst release. VOTs in Japanese are reported to be around 30 to 66 milliseconds (ms) for voiceless stops [p t k] and around −75 to −89 ms for voiced stops [b d g] (Shimizu 1999). VOT values differ across different speaking rates, positions within a word, and also across different languages, all of which cause a challenge for L2 learners.

5.1 Korean learners

One of the earliest perception experiments conducted on the L2 learners' problem with Japanese voicing contrasts was Yamada (1963) with two Korean learners of

Japanese. Yamada found that the voicing consonant distinction was difficult in word-initial position, but for [t d], word-medial was also found to be difficult. This result was replicated by many more recent studies. For example, Fukuoka (2005) conducted a perception experiment with Korean learners of Japanese at the beginning and the intermediate levels (studying Japanese for 3 and 12 months, respectively). She found that most errors occurred with voiceless stops (particularly more [t] than [p k]) at word-initial position, misperceiving them as voiced. She discussed that, because Korean does not have word-initial voiced stops and because Japanese word-initial voiceless stops have weak aspiration, Korean learners are likely to hear them as Korean tense "unaspirated" stops or lax stops (as opposed to Korean voiceless "aspirated" stops).

As for Korean learners' *production* evaluated by two NJ judges (Yokoyama 1997), word-initial voiced stops were least accurately produced, while their production in word-medial position was extremely good. Fukuoka (2007) conducted acoustic analysis on disyllables such as [papa] and [baba] produced by beginning Korean learners of Japanese. The learners' production of Japanese voiceless stops was successful and similar to that of NJs. However, their mean VOT for Japanese word-initial voiced [b] was about 25–30 ms, which was equivalent to the value of NJs' voiceless [p]. It was shown that this VOT value was similar to Korean's word initial lax stops. As for word medial position, the learners' VOT for Japanese [b] was about -30 milliseconds, which was in the NJs' acceptable range.

In Jung and Kiritani's (1998) experiment, in which Korean learners of Japanese perceived voiced vs. voiceless obstruents (stops, fricatives, and affricates), their perception was affected by the F0 onset of the following vowels. In another experiment, Jung and Kiritani edited stimuli so that the vowel onset F0 following voiced vs. voiceless obstruents were switched. The Korean learners perceived them as voiceless if followed by higher F0, and as voiced if followed by lower F0. This F0 tendency was observed in Korean speakers' *production* by Fukuoka (2008): their production of Japanese voiceless stops showed higher F0 than voiced stops, much the same way as their production of Korean aspirated and tense unaspirated stops have higher F0 than Korean lax stops. This result is consistent with an earlier observation made by Kurono (1941), who noted that Thai learners of Japanese tended to produce obstruents as voiceless when the tone or pitch was high, but as voiced when it was lower.

5.2 Beijing Mandarin and Shanghainese learners

Shanghainese has a three way stop voicing distinction: aspirated voiceless, unaspirated voiceless, and voiced, but Beijing Mandarin has two: aspirated voiceless and unaspirated voiceless. This difference in these two languages was shown to manifest differently when they learn Japanese. Fukuoka (1995) compared Beijing

Mandarin and Shanghainese speakers' perception and production of Japanese stop voicing contrasts. Most perceptual errors were made by beginning Beijing Mandarin learners (studying Japanese about 4 months) for word-medial voiceless stops which were misperceived as voiced, and this was not better even for intermediate learners (with 1.4 years of Japanese study). On the other hand, Shanghainese learners, even beginning learners, had much fewer perceptual errors. For production, Beijing Mandarin learners' VOT values for Japanese voiced stops were much greater than those of NJs in both the word-initial and word-medial positions, although the intermediate learners' values moved closer to those of the NJs. Shanghainese learners' VOTs were closer to the NJs even at the beginning level, and the intermediate learners' VOTs were even closer.

5.3 Other learners

Nishida (2003) examined Cantonese learners' production and perception of Japanese voiced and voiceless stops, and found that all learners from beginning, intermediate, to advanced levels made production and perception errors. For production, a large number of errors were made on their intended voiceless stops which were heard as voiced by NJs in both word-initial and word-medial positions, but few errors were made on the production of voiced stops. Similarly for perception, large errors were made on voiceless stops. More specifically, the learners heard the stops with large VOT (e.g., greater than 41 milliseconds for the [p]-[b] distinction) as voiceless, and with smaller VOT (which would still be a voiceless range for Japanese) as voiced.

Nishigori (1986) conducted a discrimination test between Japanese [t] and [d] and between [d] and [r] ([ɾ]) with Taiwanese learners, and found that they had difficulty discriminating these pairs, with scores lower in word-initial than word-medial position. Learners at different levels were tested, with the length of Japanese study from less than 3 months, 3 months, 6 months, and 3 years or more; their performances did not differ across levels. Many Taiwanese speakers speak Min Nan, which has three-way voicing contrasts for [p^h p b] and [k^h k g], but it has only two-way for [t^h and t]. According to Wang's (1999) anecdotal observation, Taiwanese speakers have difficulty with Japanese voicing contrasts for all places of articulation, and this difficulty may be related to nasality, though the author did not explain how nasality comes into play for voicing contrasts. This issue needs to be pursued in the future.

Minagawa (1994) examined production of Japanese voiceless and voiced stops in word-medial position across speakers of seven languages, and found that their VOTs reflected those of their native languages. For example, American English, Korean, Mandarin, and Welsh speakers' VOTs for voiceless stops were proportionately longer than those of NJ speakers. French and Finnish speakers' VOTs were

similar to those of NJs. For voiced stops, French and Finnish speakers' VOTs were negative values, which was also the case for NJ's. American English and Welsh speakers showed positive VOTs for their Japanese voiced stops, which is similar to their L1 patterns. Mandarin speakers' voiced stops were notable in their positive VOTs and extremely long stop-closure duration.

6 Theories of L2 speech acquisition

This section introduces several theories of L2 speech acquisition that have been influential in driving empirical studies world-wide. Since the 1950s (section 6.1), theories are proposed to account for ways in which speakers of any language learn any L2, and attempt to provide explanations for how and why certain difficulties occur in the process of L2 speech learning. Historically speaking, studies of Japanese as an L2 have emerged out of practical interest, such as in teaching Japanese to speakers of other languages, and have not been theoretically driven. However, a few studies have attempted to address theoretical issues in the context of Japanese as an L2, and they are introduced in sections 6.2–6.5. Where it is possible (section 6.4), ideas are suggested for possible future research that could address theoretical issues.

6.1 Earlier theories

In early years, Lado (1957) proposed that we can predict learners' difficulty if we compare and contrast phonological structures of their L1 and the target L2 (*Contrastive Analysis*). However, soon after in the 1960s, this analysis method was limited in predicting difficulties of learning specific speech sounds (e.g., Suzuki 1963; Yamada 1963). Corder (1967) then proposed the *Error Analysis*, which was a good guide for Japanese language instructors in the 1970s and 1980s (Ayusawa 1999). However, the Error Analysis was also insufficient in describing and predicting the learning process as a whole. Schachter (1974) advocated that we need to pay attention to the learner's learning entire system, including the correct usages and the usages that the learner avoids.

6.2 Interlanguage

Selinker (1972) proposed the notion of *Interlanguage*, which refers to the language system that a learner develops when he is in a process of mastering an L2 (target language). The L2 speech that the learner speaks is, in large part, not exactly the same as that spoken by a native speaker of that target language, and the Interlanguage reflects both the learner's L1 and the target language and is transient

as he progresses in his learning. Selinker proposed five major learning processes, including *language transfer* and *overgeneralization*, that predict the nature of the Interlanguage. In the area of L2 phonetics and phonology of Japanese, researchers had begun providing behavioral and observable data of Interlanguage since the 1990s (see Ayusawa 1999 for an overview). For example, Toda (1998b) measured L1 speakers' duration of single vs. geminate obstruents in Australian English across two words (e.g., *get Mary* vs. *get Tom*), and compared these measures to the duration of those in Japanese (e.g., [kate] vs. [katte]) by beginning Australian learners of Japanese. In both cases, the ratio of single and geminate obstruents was less than 1:2. This ratio is not sufficient for the Japanese target ratio of 1:2.4. This result was interpreted as negative language transfer from English to Japanese. Similarly, Fukuoka (2006) used Interlanguage Phonology (Major 1987) to explain the different degrees of success in Shanghainese and Beijing learners' acquisition of Japanese voicing contrasts as described in section 5.2. Shanghainese learners had positive, and Beijing Mandarin learners negative, transfer from their L1s on their acquisition of Japanese stop voicing at an early stage, although at a later stage, many learners were able to improve their VOT values to be more like those of NJ.

6.3 Markedness Differential Hypothesis

Eckman (1977) introduced the notion of typological markedness of speech sounds, which originated in Jakobson (1968), to account for L2 learning difficulties. A speech sound (e.g., a voiced stop) is said to be typologically marked if the presence of this sound implies the presence of another (e.g., a voiceless stop). It is proposed that unmarked sounds exist more naturally in language typology and are easier to learn. Eckman's *Markedness Differential Hypothesis* predicts that it is more difficult to learn a given L2 sound if it is a more marked sound *and* if it differs from any speech sound of the learner's L1. Yokoyama (1997) showed support of this hypothesis with Korean learners of Japanese. For example, voiced stops are "marked" *and* they do not exist in Korean word-initially, and thus, the Japanese word-initial voiced stops were predicted and shown to be most difficult for Korean learners to acquire. On the other hand, voiced stops are used word-medially in both Korean and Japanese, so the Japanese word-medial voiced stops were predicted and shown to be easier for the Korean learners to acquire (i.e., due to positive transfer from Korean).

6.4 Speech Learning Model

Flege's (1995) *Speech Learning Model (SLM)* has become an influential model that has guided a large number of empirical studies in the field of L2 acquisition since the 1990s. The model attempts to account for phonetic and phonological abilities,

eventual attainments, and limits of L2 learners across their entire lifespan, and consists of eleven postulates and hypotheses. One hypothesis is that formation of a new phonetic category in an L2 may be blocked by the mechanism of *equivalence classification* (Flege 1987a). This mechanism is responsible for an L2 phonetic category to be perceptually equated to a *similar* L1 phonetic category, and prevents the learner from pronouncing this L2 sound authentically. In contrast, an L2 phonetic category that is *new* or *different* from any L1 category is predicted to be easier to learn. Yokoyama (2000) tested this with Northern Chinese learners of Japanese, and found results that support this mechanism: Japanese voiced stops were learned without difficulty because they are substantially different from any L1 category, while Japanese voiceless stops in the word-medial position were difficult because they are similar to L1 equivalents.

Besides Yokoyama (2000), however, there have not been many studies that test the SLM in the context of Japanese as an L2. Study of Japanese as an L2 provides an opportunity to contribute greatly to testing this L2 model. For example, another hypothesis in Flege's SLM predicts that L1 and L2 phonetic categories are claimed to exist in a common phonological space, and as a consequence, the establishment of new L2 phonetic categories may affect the existing L1 categories over the long run (Flege 1987b). Given the large number of Chinese and Korean native speakers living in Japan, we could examine not only their ability to perceive and produce Japanese stop voicing contrasts (sections 5.1 and 5.2), but also their possible long-term changes in the ways they perceive and produce the stop voicing contrasts in their respective L1s. Results would be able to test the hypothesis that new L2 phonetic categories affect the existing L1 categories.

6.5 Feature Prominence Hypothesis

Highly relevant to the Japanese speech elements discussed in this chapter (sections 2.1.5 and 3.1.5) is the *Feature Prominence Hypothesis (FPH)*[1] which states that "L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner" (McAllister, Flege, and Piske 2002: 230). McAllister, Flege, and Piske (2002) examined perception of Swedish vowel quantity contrasts by three groups of listeners whose L1s differ in the degree of use of duration in their phonological distinctions: Estonian having vowel quantity distinctions, English having short and long vowels that simultaneously differ in their formant frequencies (e.g., English [i]-[ɪ]), and Spanish having no phonological quantity distinctions for which

1 A contrasting hypothesis worth mentioning is Bohn's (1995) *Desensitization Hypothesis*, stating that L2 learners resort to durational information in attempting to hear L2 vowel contrasts if they cannot hear (i.e., are desensitized to) the contrasting spectral differences of those vowels. See Cebrian (2006) and Kondaurova and Francis (2008) for supporting evidence.

duration is a major perceptual cue. The results supported the FPH: Estonian participants performed at the level of native Swedish speakers, whereas English participants performed less well than the Estonian participants, but better than Spanish participants. Hirata and Ueyama's (2009) study also gives partial support to this hypothesis. Italian, but not English, has consonant length contrasts similar to those in Japanese. Native Italians speakers outperformed native speakers of English in detecting the presence of geminate consonants as opposed to single consonants in sentences. (This Italian speakers' advantage was not found in an isolated word context, however, which is why this study provides only *partial* support for the hypothesis.) Furthermore, neither Italian nor English has vowel length contrasts as Japanese does, and the two groups of subjects showed no difference in their abilities to detect short and long vowel contrasts, which is also in line with the FPH.

The FPH can be extended to test whether the L1 prosodic feature of pitch accent or tones affects L2 learning. Masuko and Kiritani (1990) is one of the earliest studies to examine effects of native languages on the learners' perception of Japanese lexical pitch accent. Although this study was conducted before the formation of the FPH, results of this study support this hypothesis. Participants of tonal languages (Chinese and Thai) identified the Japanese lexical pitch accent patterns better than participants of non-tonal languages (Indonesian and Korean). A study by Wayland and Li (2008) was also supportive of this hypothesis with regard to L2 learning of Thai tones, showing that native Chinese (tone language) speakers outperformed English (non-tone language) speakers learning to identify and discriminate Thai tones.

Some studies do not support the FPH, however. As an example of duration as a phonological feature, Arabic has a vowel length distinction that is similar to Japanese, but Australian English does not. Tsukada (2011b) found that, despite the FPH's prediction, native Arabic speakers were not any better perceiving Japanese vowel length distinction than native speakers of Australian English. Results of Minagawa and Kiritani (1996) and Minagawa (1996) seem to be also inconsistent with the FPH: Although none of Spanish, English, Korean, Thai, and Chinese has consonant length distinction, Spanish and English speakers did better than speakers of the other languages in perceiving Japanese single and geminate consonant distinctions (section 2.1.5). This result cannot be explained by the FPH. In addition, perceptual learning of Cantonese lexical tones by native speakers of English and Mandarin (non-tonal vs. tonal languages) did not show a clear advantage for Mandarin speakers (Francis et al. 2008). Furthermore, So (2005) found that Japanese speakers (which is not a tonal language but has phonemic pitch contrasts) did slightly better than Cantonese speakers (a tonal language) in the identification of Mandarin tones after perceptual training. So (2005) suggested that the Cantonese tonal system hindered their learning of Mandarin tones, and used Best's (1995) Perceptual Assimilation Model to account for this result.

In summary, there are almost equal numbers of studies that support and do not support the FPH, and thus future research is necessary to come to a clear con-

clusion. It would be useful to pursue this line of research more in the context of Japanese as an L2 because Japanese has the prosodic feature of duration and lexical pitch accent, both of which cause notable difficulties for non-native learners as described in sections 2–4.

6.6 Summary

The models and hypotheses introduced in this section are valuable because they attempt to predict L2 speech learning patterns beyond predictions from a particular L1 towards learning of a particular L2. The more generalizations we can make regarding the learners' initial states and ultimate attainments, the less we need to describe and predict L2 acquisition in ad hoc ways for every combination of L1 and L2. These models and hypotheses will continue to provide directions for empirical research, and empirical data will help to advance the theoretical aspects of L2 speech acquisition. Examination of Japanese as learned by speakers of other languages has potential value in elucidating many issues addressed in these theoretical pursuits.

7 Training and technology

For both theoretical and practical interests, abundant research has been conducted to train non-native learners of Japanese to perceive and produce the challenging speech elements described in sections 2–5. A variety of *perception* training on Japanese length contrasts and lexical pitch accent are reviewed in section 7.1, where several factors are found to play a significant role in learning, such as speaking rates of stimuli and visual information accompanying auditory stimuli. In section 7.2, different types of *production* training are reviewed, and the efficacy of a variety of feedback given to learners' production is examined.

7.1 Perception training

7.1.1 Auditory training on Japanese length contrasts

One of the earliest studies that scientifically examined effects of perceptual training on non-native speakers' perceptual learning was Yamada, Yamada, and Strange (1995). Their perception training included triplets of disyllables such as [kaka, ka:ka, kakka] and [sasa, sa:sa, sassa] spoken by three NJs, and asked learners to identify whether the word contained a long vowel, a geminate obstruent, or neither. They found that intensive training with eight sessions each with 270 stimuli enabled the learners to improve their perception performance.

Hirata (2004b) went a step further, investigating effects of training with sentences as compared to that with isolated words. The study aimed at enabling perceptual learning that is useful for perceiving words of varied length in sentences. The training provided a variety of words of one- to six-mora length, including short and long vowels and single and geminate obstruents (but not necessarily in minimal pairs), and asked learners to count the number of moras in the words. For example, words such as [ʃisso] 'simple' (3 moras), [se:bɯtsɯgakɯ] 'biology' (6 moras), [tasse:] 'accomplishment' (4 moras), and [tsɯbo] 'a jar' (2 moras) were presented in a random order, and participants were trained to count the number of moras in these words. Native English speakers with no knowledge of Japanese initially counted the number of syllables (2, 5, 2, and 2 syllables, respectively, in the above examples), but were unable to detect the long vowels and geminate obstruents that add a mora to the number of syllables. However, within ten training sessions each with only 60 trials, their perception significantly improved, compared to the control group that did not participate in any training. Furthermore, the group that heard these words in a variety of carrier sentences improved in both contexts of isolated words and words-in-sentences, whereas the group that heard the words only in isolation showed less generalization from the isolated word context to the words-in-sentence context. This study suggests that it may be more beneficial to train learners in sentence contexts if their ultimate goal is to be able to hear these difficult sound distinctions in fluent speech.

A similar finding was obtained in Sonu et al. (2011b) with regard to effects of word- vs. sentence-training. Sonu et al. trained Korean learners to identify short and long vowels in Japanese using a two-alternative forced choice identification task. The word-training and the sentence-training groups did not differ in the overall amount of perceptual improvement, but the sentence-training group, more than the word-training group, showed an ability to generalize for untrained contexts.

7.1.2 Speaking rate as a factor for length contrast training

Tajima et al. (2008) examined how training can improve learners' abilities to distinguish a variety of length contrasts in Japanese. Native speakers of Canadian English with no knowledge of Japanese were trained to identify Japanese short and long vowels in words spoken in isolation at a normal speaking rate. The trained group was tested in their ability to perceive these vowel length distinctions (e.g., [kaze] 'wind' vs. [kaze:] 'taxation'), as well as their ability to generalize their learning to other length distinctions such as obstruent, nasal (e.g., [tanin] 'other people' vs. [tannin] 'a person in charge'), and palatal pairs (e.g., [kjakɯ] 'a guest' vs. [kijakɯ] 'regulations'). Tajima et al. found that the trained group's overall improvement did not significantly differ from that of the control group who did not participate in training but only took the test twice. However, the trained group improved significantly more than the control group on the vowel length pairs. The results indicate

that learners do not transfer their learning of vowel length to that of consonant length. Another interesting question addressed in this study concerned the extent to which training in one context and with one speaking rate generalizes to another context and to other speaking rates. The test included words in isolation and in carrier sentences, spoken at three speaking rates, while the training was only with isolated words, spoken at a normal speaking rate. Tajima et al. found that the ability gained through isolated-word training did not transfer well to the sentence context and to different speaking rates.

Hirata, Whitehurst, and Cullings (2007) tested Pisoni and Lively's (1995) High Phonetic Variability Hypothesis. This hypothesis states that the more phonetically and acoustically varied speech materials learners receive, the more robustly they learn to form new L2 categories and perceive difficult L2 phonemic distinctions. Hirata, Whitehurst, and Cullings examined whether the learners' abilities to perceive Japanese vowel length contrasts improve with only slow-rate materials (slow-only training), only with fast-rate materials (fast-only training), or with both slow and fast speech materials (slow-fast training). The experimental task was to identify whether the second vowel of various target disyllabic words spoken in a carrier sentence was short or long, e.g., [ise] (name of a place) vs. [ise:] 'opposite gender'. The group of native English speakers with no knowledge of Japanese who participated in the slow-fast training improved significantly more than a control group who did not participate in any training. The group that received the slow-only training improved, but the amount of their improvement was only marginally more than that of the control group. Finally, the group that received the fast-rate training improved the least and the amount of their improvement did not significantly differ from that of the control group. Pisoni and Lively's High Phonetic Variability Hypothesis was supported in a sense that the slow-fast training was more effective for non-native learners' perceptual learning than the slow-only training. As for the result that the slow-only training was more effective than the fast-only training, Hirata, Whitehurst, and Cullings interpreted it as also supporting the High Phonetic Variability Hypothesis in the sense that slow speech is generally more varied than fast speech in terms of absolute duration (Hirata 2004a). This study also shows that the speaking rate that learners are exposed to does affect their perceptual learning.

It is interesting to note that in Sonu et al. (2009), Korean learners of Japanese who were trained for single and geminate consonants did not show a distinct advantage of three-rate training over one-rate training when stimuli were isolated words. There are a number of differences in experimental settings of Hirata, Whitehurst, and Cullings (2007) and Sonu et al. (2009), and it would be interesting to narrow down specific factors that contributed to these different results.

7.1.3 Auditory training on pitch accent

Compared to perceptual training on length contrasts, there are fewer studies examining effects of perceptual training on acquisition of lexical pitch accent in Japanese

(but see section 7.2.1 for *production* training). Shport (2011) trained native English speakers to perceive Japanese lexical pitch accent in two-mora words followed by a particle: H'L+L, LH'+L, and LH+H, e.g., [ɯmi] 'sea', 'pus', and 'giving birth' in these three accent patterns, respectively. Learners were trained to hear these pitch patterns in various carrier sentences and to choose one of the three alternatives in a one-hour training session. Shport (2011) divided participants into low-score vs. high-score groups based on their pretest scores for both control and training groups. The low-score trained group improved significantly more than the low-score control group, showing a small (though not statistically robust) effect of training. The high-score trained group already had high scores similar to NJs at the pretest, and thus their improvement after training did not differ from the high-score control group. With regard to the three pitch patterns, the scores on the words with LH+H were lowest and showed least improvement after training. Shport's experimental design attempted to test Pisoni and Lively's High Phonetic Variability Hypothesis, using various target words, varied contexts with different carrier sentences, and different speakers. It is notable that the trained learners generalized their learning to untrained word pairs that were spoken by the familiar speaker who appeared in training.

7.1.4 Auditory training combined with visual information

Motohashi-Saigo and Hardison (2009) conducted an experiment with beginning native English learners of Japanese to compare effectiveness of two training methods: one with audio materials only and another with waveform displays on which a cursor moved along with the auditory input. Target speech sounds used in training were 120 Japanese real and nonsense words with singleton and geminate obstruents for each of ten sessions. Learners' abilities to perceive and produce these words were tested before and after training, and their scores were compared between the audio-only and the audio-visual groups. Results showed a distinct advantage of the auditory-visual training for both perception and production.

Hirata and Kelly (2010) examined relative effects of two pieces of visual information: mouth movement that goes along with an NJ speaking sentences and hand gestures that beat the rhythm of a short and a long vowel. Four groups of native English speakers received identical auditory stimuli: ten target nonsense disyllables that differed in the length of the second vowels, e.g., [mimi] and [mimiː], spoken in a carrier sentence. The only difference among the four groups was the visual information used during training: (1) Audio-only training with still images of native Japanese speakers, (2) Audio-Mouth training in which the speakers were moving their mouth along with the audio, (3) Audio-Hands training in which the speakers' mouth movement was not visible but the hand gesture showed the beats of the words, and (4) Audio-Mouth-Hands training in which participants were able to see both the mouth and hand movements along with the audio. Before and after the training,

the ability to identify the short and long vowels in words that were not used in training was tested without any visual information. Hirata and Kelly found that the Audio-Mouth group improved significantly more than the Audio-only group, showing a distinct effect of seeing mouth movements in auditory learning of Japanese vowel length contrasts. However, the Audio-Hands and the Audio-Mouth-Hands groups did not improve more than the Audio-only group, indicating that seeing the hand gesture did not have a distinct effect on auditory learning. It is notable that having all of the information, i.e., audio, mouth, and hands, cancelled out the positive effect of mouth movements, which Hirata and Kelly discussed as a possible cognitive overload or visual distraction.

7.2 Production training

7.2.1 Production training with visual feedback

Masuko, Imagawa, and Kiritani (1989), Saita et al. (1992) and Landahl et al. (1992) were the early pioneers who explored innovative production training that aimed at improving L2 learners' pronunciation of Japanese. Masuko, Imagawa, and Kiritani (1989) explored developing pronunciation training using a personal computer that exhibits real-time F0 contours of speech on the screen. With this training program, learners were able not only to listen to NJ model speech and their own, but also to see and compare the model's and the learners' F0 contours. Although no systematic experiment was conducted, Masuko, Imagawa, and Kiritani qualitatively described this innovative use of technology for training on Japanese pitch accent as a future possibility. Saita et al. (1992) also provided an important step towards the use of technology in training production of geminates and long vowels in Japanese. Waveforms of NJ model speech and those of learners are displayed on a computer screen so that learners are able to notice how accurately they are pronouncing targets. No experimental data were provided, but one case of a learner was given, who first produced a word without an appropriate geminate but was able to finally produce it accurately.

Landahl et al. (1992), Landahl and Ziolkowski (1995), and Ziolkowski and Landahl (1995) investigated experimentally effects of visual F0 feedback on L2 learners' production of Japanese vowel and consonant length contrasts and lexical pitch accent. They used a method similar to Masuko, Imagawa, and Kiritani (1989) that displayed F0 contours of Japanese word pairs produced by NJs and learners. Landahl's group compared this F0 display method with a traditional listen-and-repeat method and a practice-with-a-tutor method, each for a length of an hour. They found that while all three methods helped the learners improve their productions, the tutor method was most effective. While no singular advantage of the F0 display method was found over the other methods, at least for consonant length contrasts, the F0 display

method was as effective as the tutor method, whereas the listen-and-repeat method tended to yield productions that were exaggerated in durational distinctions.

Hirata (1999, 2004c) developed a production training program using real-time F0 contours of learners' productions that was compared with those of NJ models. In order for the learner to understand the break-downs of F0 contours, Hirata used prosody graphs that showed schematic moras and pitch height of the target words, phrases, and sentences. The training consisted of ten 30-minute sessions each of which was accompanied by written instructions and explanations of how the pitch accent and sentence intonation should be produced. The trained group improved both perception and production of pitch accent for both isolated words and for sentences.

Hirano-Cook (2011) investigated whether a series of instructions (30 minutes × 6 sessions) could assist learners in improving their perception and production of Japanese. Hirano-Cook's approach was based on various teaching techniques (e.g., Lee and VanPatten 2003; González-Bueno 2005) that included lectures on Japanese pitch accent, rhythm, and intonation, class exercises including peer learning activities, visualization of pitch accent in fundamental frequency contours and schematic graphs, and multimodal practice with hands and neck movements. The group that underwent this series of instructions and practices improved in their perception significantly more than the control group that did not undergo this series. As for the production performance, both the trained group and the control group showed improvement, and thus the improvement could not be attributed to the experimental instructions per se. However, when the participants did not receive written accent symbols to guide their production, the trained group did significantly better than the control group. Hirano-Cook's study suggests that this kind of phonetic learning in perception and production *can* take place in realistic class settings by raising phonetic awareness and fostering self-monitoring skills, instead of intense and repetitive perception or production training typically conducted in research laboratories.

7.2.2 Automatic evaluation of learners' speech

Kawai and Hirose (2000) developed a computer-assisted language learning system that aimed at training learners to produce vowel, obstruent, and nasal length distinctions, e.g., [kado] 'a corner' vs. [ka:do] 'a card,' [hata] 'a flag' vs. [hatta] 'posted', and [kona] 'powder' vs. [konna] 'this kind of'. The system aligned segments of the words that learners produced to those of NJ models and automatically evaluated duration of the target segments (e.g., duration of [a(:)] in [ka(:)do]), based on the previous results on how NJs perceived these segments. Then the system gave feedback that was helpful for the learners, e.g., "Your kado can be understood by 100% of native speakers, but your kaado can be understood by only 10%." (Kawai and Hirose 2000: 135–136). Kawai and Hirose conducted an evaluation experiment and

found that learners' production accuracy improved. Strictly speaking, in order to test the true effects of training it would be necessary to compare this result with the result of a control group which did not participate in the automated training but produced the same speech twice. However, this was one of the earliest studies that developed a system to automatically evaluate non-native learners' production of Japanese with immediate feedback without a human instructor.

Along the same line, Tsurutani (2008) developed computer-assisted pronunciation practice software in which learners' recordings of Japanese sentences were automatically evaluated for accuracy using an automatic speech recognition system. With this software, learners immediately receive a score for every phrase of a sentence, and receive feedback on how they mispronounced key elements such as vowel length, consonant length, and segmental substitutions. Twenty-three students practiced with this software twice, once at the beginning and once at the end of a semester. Results showed that ten students increased their scores the second time (at the end of the semester), while five students decreased their scores. The precise effectiveness of this software is yet to be determined. However, this is an important effort for the practical use of technology in effective learning of L2 speech sounds.

7.2.3 Verbo-Tonal (VT) method

The *Verbo-Tonal* or *VT method* has a long history since the 1960s in the field of L2 language teaching and learning in Europe, the theoretical base of which was originally developed by Petar Guberina (Kawaguchi 2008). It utilizes tension and relaxation of the body and hands to help learners acquire authentic pronunciation. The VT method has widely been used in France and has been incorporated in various teaching materials for French as an L2. The method was introduced to Japan in the 1980s (e.g., Roberge, Kimura, and Kawaguchi 1996), and since then, a variety of attempts have been made in developing methods to specifically teach Japanese as an L2. Kawaguchi (2008) gives an overview of how this method works, as well as complementary use of this method along with other pedagogical approaches such as "Communicative Approach".

Using the principle of muscle tension and relaxation in the VT method, Fukuoka (1996) trained five Beijing learners of Japanese to produce Japanese voiced stops, which are known to be difficult for Mandarin native speakers. Fukuoka conducted a 40-minute session of training in which the learners practiced swinging both arms down, slouching, and slowly lowering and loosening hard fists as they produced Japanese voiced stops [b d g]. Before this training, the learners' VOT values for the voiced stops were positive, but after training, many of them became negative values as they should be. In addition, their closure duration was too long before training, but it became shorter after training. NJs' perception of these produced tokens also improved. Similarly, Ota (2003) and Jiang (2007) reported that Chinese and Korean

learners of Japanese, respectively, improved in their production of single and geminate obstruents after using the VT methods.

However, none of these studies compared these learners' performances with a control group that went through a traditional listen-and-repeat training for the exact same amount of practice time, and it is not clear whether it is this specific method that was effective or whether any training could yield similar production improvement. To affirm that the observed improvement was not due to an effect of mere task repetitions, we would need to compare the target learners of the VT method with a control group that takes the test twice with no other training. Thus, unique effects of this method have not yet been shown clearly from these studies. Kawaguchi (2008) points out the need for studies that would scientifically evaluate effects of the VT method on phonetic and phonological learning of Japanese as an L2. The multimodal aspect of this method and the improvement that many Japanese instructors have observed are promising, and it is hoped that more controlled scientific studies would prove true efficacy.

8 Future studies

Although this chapter does not include a number of topics, the following deserve attention in the field of L2 phonetics and phonology in the future. Japanese segments that have been documented to be difficult for L2 learners include Japanese [s ts dz] vs. [ʃ tʃ dʒ] consonant contrasts for learners whose native languages are, e.g., Indonesian (Sato 1986), Russian (Funatsu and Kiritani 2000), and Korean (Ho 2008; Marushima et al. 2011). Others include Taiwanese learners' distinctions between [d] and [r] ([ɾ]) (Liu 2002), their production of devoiced vowels (Hung 2003), Korean learners' distinction of fricative and affricates (e.g., [s] and [ts], Yamakawa and Amano 2010), and learners' general problems in producing the moraic nasal (Imada 1973). For suprasegmentals, sentential intonation has been studied but needs further research (Nozawa and Shigematsu 2003; Nozawa and Shigematsu 2006; Eda, Naito, and Hirano 2009).

Another topic in phonological L2 acquisition to be investigated in the future is L2 learners' latent abilities to decode newly encountered Japanese loan words into the original language, e.g., decoding [irɯmine:ʃon] as 'illumination', or the other way around, e.g., converting 'athletic' to [asɯretʃikkɯ] with appropriate moras. In addition, the phonological ability to understand abbreviated words, e.g., [pasokon] 'personal computer', and the ability to combine two Japanese words without having heard them before into one with appropriate *rendaku*.

Recent studies have also asked the question of whether and how L2 learners' abilities to perceive and produce certain elements of speech sounds relate to their vocabulary size (Bundgaard-Nielsen, Best, and Tyler 2011) or to extensive experience in music (Sadakata and Sekiyama 2011; Nakata 2002). Since the beginning of the 21st

century, brain research has begun examining how L2 learning manifests itself in the brain (Menning et al. 2002; Minagawa-Kawai, Mori, and Sato 2005; Hisagi et al. 2010; Wu, Tu, and Wang 2011). These areas of research are indicative of the fact that L2 research is extremely interdisciplinary.

On the topics covered in this chapter, readers are encouraged to go to the cited references for specific issues to explore, as well as scientifically replicating their findings. As for future directions, researchers in the field of L2 phonetics and phonology of Japanese have accumulated abundant empirical data and should be able to contribute to evaluation, revision, and proposal of L2 speech acquisition theories. Some outstanding questions are the following. What are the implications of the accumulated data for L2 theories? In what ways do the extant theories account for the existing findings in Japanese as an L2, in what ways are the extant theories insufficient, and in what ways do they require revisions or new proposals? What are common vs. language-specific phenomena related to processing of spoken Japanese and other languages as L2s? As much as the extant theories can provide guidance to our empirical research directions, the obtained and accumulated data of Japanese should be able to feed back to development of L2 speech acquisition theories.

Research in language-training, which has been one of the richest areas in the field of L2 speech acquisition of Japanese, will also benefit from having more interaction with L2 theories. For example, it is useful in the future to think more about how we can predict and determine learners' eventual attainment, maximum ability, and limit in their L2 speech learning (cf. Lenneberg 1967) when learners of a given native language are given certain types of input, and how the extant theories are sufficient or insufficient in predicting those. An answer to the question of whether learners can ultimately overcome difficulty in acquiring certain speech sounds would provide not only theoretical but also practical benefits. If it is scientifically proven that adults are limited in how well they can acquire certain speech elements, instructors should take this into account in their curricula (Ayusawa 2001), and we should promote understanding of this fact in the society. In the meantime, however, it is worth pointing out that there is much to be explored in the role of multimodal input for auditory learning of L2 speech, as it was reviewed in section 7.

For the future, collaboration between L2 researchers and instructors of Japanese as an L2 will continue to be essential. L2 researchers and speech scientists should not quickly discard anecdotal and descriptive observations given by instructors and learners just because they lack scientific evidence coming from well-controlled experiments. Language instructors often have insightful intuitions that provide good directions for scientific investigations. Language instructors should also be encouraged to think how scientific findings can be used to improve practical teaching methods or to develop instructional materials, instead of considering scientific findings too narrowly focused to have practical validity. The two camps can feed each other to advance the field of L2 phonetics and phonology. Furthermore, L2 phonetics and

phonology intersect with a wide variety of fields including psycholinguistics, psychology, neuroscience, physics, and engineering, to name a few, and further collaboration with these fields will help in attaining our goals.

Acknowledgments

This work was partially supported by Colgate University's Research Council.

References

Ayusawa, Takako. 1999. Interlanguage studies: Japanese learners' sound systems. *Journal of the Phonetic Society of Japan* 3(3). 4–12.

Ayusawa, Takako. 2001. Phonetic studies for Japanese language education. *Journal of the Phonetic Society of Japan* 5(1). 71–74.

Ayusawa, Takako. 2003. Acquisition of Japanese accent and intonation by foreign learners. *Journal of the Phonetic Society of Japan* 7(2). 47–58.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Timonium, Maryland: York Press.

Bohn, Ocke-Schwen. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 279–304. Timonium, Maryland: York Press.

Bundgaard-Nielsen, Rikke L., Catherine T. Best and Michael D. Tyler. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33. 433–461.

Cebrian, Juli. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34. 372–387.

Corder, S. P. 1967. The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching* 5(4). 161–170.

Eda, Sanae, Yuka Naito and Erika Hirano. 2009. Gakushūsha ni yoru intonēshon chikaku to imi rikai no sutoratejī: Onsei kyōiku e no ōyō to teigen [Strategic use of intonational cues by learners of Japanese: Application for teaching Japanese as a foreign language]. *Nihongo Kyōiku* 143. 48–59.

Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27(2). 315–330.

Enomoto, Kayoko. 1992. Interlanguage phonology: The perceptual development of durational contrasts by English-speaking learners of Japanese. *Edinburgh Working Papers in Applied Linguistics* 3. 25–35.

Flege, James E. 1987a. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47–65.

Flege, James E. 1987b. A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8. 162–177.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. Timonium, Maryland: York Press.

Francis, Alexander L., Valter Ciocca, Lian Ma and Kimberly Fenn. 2008. Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers. *Journal of Phonetics* 36(2). 268–294.

Fujisaki, Hiroya, Kimie Nakamura and Toshiaki Imoto. 1975. Auditory perception of duration of speech and non-speech stimuli. In Gunnar Fant and Marcel A. A. Tatham (eds.), *Auditory analysis and perception of speech*, 197–219. London: Academic Press.

Fukuoka, Masako. 1995. Pekingo bogo washa to shanhaigo bogo washa o taishō to shita nihongo no yūsei musei haretsuon no ōdanteki shūtoku kenkyū [A longitudinal study in the acquisition of voiced and voiceless stops by native speakers of Beijing Mandarin and Shanghainese]. *Gengo Bunka to Nihongo Kyōiku* 9. 201–215.

Fukuoka, Masako. 1996. Nihongo no yūsei haretsuon no shūtokujō no mondaiten to VT-hō o tsukatta hatsuon kyōsei [Difficulty in acquisition of Japanese voiced stops and pronunciation training using the VT method]. *Nihongo Kyōiku Kenkyū* 32. 97–110.

Fukuoka, Masako. 2005. A longitudinal study of the acquisition of Japanese plosives by Korean learners of Japanese in perception. *Mie University Bulletin of Center for International Students* 7. 25–34.

Fukuoka, Masako. 2006. Interference and developmental process: Creating an Interlanguage through acquisition of plosives. *Mie University Bulletin of Center for International Students* 8. 15–28.

Fukuoka, Masako. 2007. Plosive acquisition study of Koreans studying Japanese: The mechanism of perception and pronunciation. *Mie University Bulletin of Center for International Students* 2. 13–22.

Fukuoka, Masako. 2008. Pitch pattern of words containing plosives in word initials pronounced by Korean Japanese language learners. *Mie University Bulletin of Center for International Students* 3. 45–59.

Funatsu, Seiya and Shigeru Kiritani. 2000. Effects of following vowel on perception of second language fricatives: Native language interference in Russian learners of Japanese. *Journal of the Phonetic Society of Japan* 4(2). 72–80.

González-Bueno, Manuela. 2005. Articulatory difficulties in the acquisition of Spanish /r/ in a bilingual context. In James Cohen, Kara T. McAlister, Kellie Rolstad and Jeff MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism*, 914–934. Somerville, MA: Cascadilla Press.

Han, Mieko. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49. 102–127.

Hardison, Debra M. and Miki Motohashi. 2010. Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics* 31. 81–99.

Hasegawa, Tsuneo. 1977. Kantongo no on'in kōzō to kantongo washa no nihongo hatsuon [Phonological structures of Cantonese and native Cantonese speakers' pronunciation of Japanese]. *Nihongo to Nihongo Kyōiku* 6. 13–48.

Hirano-Cook, Erika. 2011. *Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception, and production*. Lawrence, KS: University of Kansas dissertation.

Hirata, Yukari. 1990a. Perception of geminated stops in Japanese word and sentence levels. *The Bulletin* (The Phonetic Society of Japan) 194. 23–28.

Hirata, Yukari. 1990b. Perception of geminated stops in Japanese word and sentence levels by English-speaking learners of Japanese language. *The Bulletin* (The Phonetic Society of Japan) 195. 4–10.

Hirata, Yukari. 1999. *Acquisition of Japanese rhythm and pitch accent by English native speakers*. Chicago, IL: The University of Chicago dissertation.

Hirata, Yukari. 2004a. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32. 565–589.

Hirata, Yukari. 2004b. Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America* 116. 2384–2394.

Hirata, Yukari. 2004c. Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning* 17. 357–376.

Hirata, Yukari. 2009. Factors affecting the perception of Japanese moraic rhythm by second language learners. *Journal of the Phonetic Society of Japan* 13(3). 33–43.

Hirata, Yukari and Spencer D. Kelly. 2010. Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research* 53(2). 298–310.

Hirata, Yukari and Steven G. Lambacher. 2004. Role of word-external contexts in native speakers' identification of vowel length in Japanese. *Phonetica* 61. 177–200.

Hirata, Yukari and Ian McNally. 2010. Japanese language learners' production of short/long vowel and single/geminate consonant contrasts: A longitudinal development. *Journal of the Acoustical Society of America* 127(3). 1852.

Hirata, Yukari and Motoko Ueyama. 2009. Does the native language use of duration affect the perception of non-native length contrasts? *Journal of the Acoustical Society of America* 125(4). 2771.

Hirata, Yukari, Elizabeth Whitehurst and Emily Cullings. 2007. Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *Journal of the Acoustical Society of America* 121(6). 3837–3845.

Hirato, Nobuo and Shinichiro Watanabe. 1987. Sokuon no chikaku to kōzoku boin no jizoku jikan to no kankei [The relationship between the perception of moraic obstruents and the duration of the following vowels]. *Onsei Gengo* II. 99–106.

Hisagi, Miwako, Valerie L. Shafer, Winifred Strange and Elyse S. Sussman. 2010. Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research* 1360. 89–105.

Ho, Sun-Jo. 2008. Kankokugo o bogo to suru jōkyū nihongo gakushūsha ni yoru zagyōon no shūtoku [Acquisition of za-line sounds by Korea native speakers learning Japanese at an advanced level]. In Takako Toda (ed.), *Japanese language education and speech communication*, 163–182. Tokyo: Kurosio.

Horigome, Mio. 1999. Kankokugo washa to nihongo washa ni okeru sokuon chikaku no tegakari ni tsuite [On the perceptual cues to moraic obstruents by native speakers of Korean and Japanese]. *Kotoba no Kagaku [Studia Linguistica]* 12. 275–286.

Hung, Hsin-Yi. 2003. Taiwan no nihongo gakushūsha ni okeru boin museika [Vowel devoicing by Taiwanese learners of Japanese]. In Hiroaki Endo, Xin-Yi Hong, Hisaki Satoi, Satoshi Yabuuchi and Shinichiro Watanabe (eds.), *Onsei gengo no kenkyū*, 7–18. Osaka: Osaka Daigaku Gengo Bunka Bu.

Hung, Hsin-Yi. 2012. Perceptions of Japanese geminate stops among Taiwanese learners of Japanese. *Journal of the Phonetic Society of Japan* 16(2). 15–27.

Imada, Shigeko. 1973. Hatsuon no hatsuon o megutte [On the pronunciation of the moraic nasal]. *Nihongo Kyōiku* 20. 73–82.

Irie, Michio. 1941. Watashi no nihongo kyōjuhō benkyō [My study of Japanese language pedagogy]. *Kotoba* 12. 27–34.

Jakobson, Roman. 1968. *Child language, aphasia and phonological universals*. The Hague: Mouton.

Jia, Haiping, Hiroki Mori and Hideki Kasuya. 2005. Effects of speaking rate on mora timing organization in producing Japanese contrastive geminate/single consonants and long/short vowels by native and Chinese speakers. *Acoustical Science and Technology* 26(6). 533–536.

Jiang, Lian-Hua. 2007. Kankokujin nihongo gakushūsha ni okeru tokushu onso no sokuon no chikaku to seisei ni kansuru kenkyū [Research in perception and production of moraic obstruents by Korean learners of Japanese]. *Takushoku Daigaku Gengo Kyōiku Kenkyū* 7. 91–103.

Jung, Eun-Jung and Shigeru Kiritani. 1998. Pitch patterns in Japanese and their influence on the distinction between voiced and voiceless sounds: A comparison between Korean and Japanese speakers. *Journal of the Phonetic Society of Japan* 2(2). 64–70.

Kanda, Toshiko and Tomoko Uozumi. 1995. Rizumu kyōiku no kōka ni tsuite [On effects of rhythm training]. *Nagoya Daigaku Nihongo Nihon Bunka Ronshū* 3. 61–73.

Kashima, Tanomu. 2003. Acquisition of segments in Japanese by foreign learners. *Journal of the Phonetic Society of Japan* 7(2). 59–69.

Kawagoe, Itsue. this volume. Chapter 2: The phonology of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawaguchi, Yoshikazu. 2008. VT (Verubo-tonaru) hō ni yoru nihongo onsei shidō. In Takako Toda (ed.), *Japanese language education and speech communication*, 117–138. Tokyo: Kurosio.

Kawahara, Shigeto. this volume. Chapter 1: The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawahara, Shigeto. this volume. Chapter 11: The phonology of Japanese accent. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Kawai, Goh and Keikichi Hirose. 2000. Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication* 30. 131–143.

Kawakami, Shin. 1984. The metrical distinction of Japanese words and aid of teaching Japanese. *The Bulletin* (The Phonetic Society of Japan) 175. 7–9.

Kindaichi, Haruhiko. 1963. Hatsuon kara mita nihongo [Phonetic aspects of Japanese]. *Nihongo Kyōiku* 3. 2–18.

Kondaurova, Maria V. and Alexander L. Francis. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America* 124(6). 3959–3971.

Kurihara, Michiyo. 2007. The preliminary studies on perception of vowel length by learners of Japanese: In the case of Finnish, Chinese and Korean speaking learners of Japanese. *Nihongo Kyōiku Hōhō Kenkyūkaishi* 14(1). 62–63.

Kurono, Masaichi. 1941. Nihongo kyōjujō no onsei mondai [Phonetic issues in teaching Japanese]. *Kotoba* 7. 22–34.

Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: The University of Michigan Press.

Landahl, Karen and Michael Ziolkowski. 1995. Discovering phonetic units: Is a picture worth a thousand words? *Proceedings of the Annual Meeting of the Chicago Linguistic Society 31-I*, 294–316.

Landahl, Karen, Michael Ziolkowski, Mayumi Usami and Brenda Tunnock. 1992. Interactive articulation: Improving accent through visual feedback. *Proceedings of the Second International Conference on Foreign Language Education and Technology*, 283–292.

Lee, James F. and Bill VanPatten. 2003. *Making communicative language teaching happen*. New York: McGraw-Hill, Inc.

Lenneberg, Eric H. 1967. *Biological foundations of language*. New York: John Wiley and Sons, Inc.

Liu, Chiu-Yen. 2002. Recognition of stop /d/ and flap /r/ by Japanese native speakers and Taiwanese learners of Japanese: Using closure duration, spike strength and spike (transition) duration as variables. *Journal of the Phonetic Society of Japan* 6(3). 69–78.

Liu, Shu-Yuan. 1984. Chūgokujin gakushūsha ni yoku mirareru hatsuonjō no ayamari to sono kyōsei hōhō [Pronunciation errors and their correction methods for Chinese learners of Japanese]. *Nihongo Kyōiku* 53. 93–101.

Major, Roy C. 1987. A model for interlanguage phonology. In Georgette Ioup and Steven Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, 101–124. Cambridge, MA: Newbury House.

Marushima, Ayumi, Mai Kirikishi, Takashi Ninomiya, Kazuki Watanabe, Yurie Hayakawa and Takahiro Fukumori. 2011. Distinction between [dz] and [ʥ] among Korean learners of Japanese. *Research in Experimental Phonetics and Linguistics* 3. 12–29.

Masuda, Kyoko. 2009. Influence of L1 on the acquisition of Japanese mora timing by native speakers of English and Korean. *Nihongo Kyōiku* April. 3–13.

Masuda, Kyoko and Rachel Hayes-Harb. 2005. Eigo bogo washa ni yoru sokuon seisei no sutoratejī [Native English speakers' strategies of producing moraic obstruents]. In Masahiko Minami (ed.), *Gengogaku to nihongo kyōiku IV: New directions in applied linguistics of Japanese*, 259–270. Tokyo: Kurosio.

Masuda, Kyoko and Rachel Hayes-Harb. 2007. Lexical development of Korean learners of Japanese: A case study of Japanese consonant quantity. In Masahiko Minami (ed.), *Applying theory and research to learning Japanese as a second language*, 40–57. Newcastle, UK: Cambridge Scholars Publishing.

Masuko, Yukie, Hiroshi Imagawa and Shigeru Kiritani. 1989. Japanese pronunciation-training using a personal computer. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 23. 65–74.

Masuko, Yukie and Shigeru Kiritani. 1990. Perception of mora sounds in Japanese by non-native speakers of Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 24. 113–120.

Matsuzaki, Hiroshi. 1999. Phonetic education of Japanese for Korean speakers. *Journal of the Phonetic Society of Japan* 3(3). 26–35.

McAllister, Robert, James E. Flege and Thorsten Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

Menning, Hans, Satoshi Imaizumi, Pienie Zwitserlood and Christo Pantev. 2002. Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory* 9(5). 253–267.

Min, Kwang Joon. 1987. Kankokujin no nihongo no sokuon no chikaku ni tsuite [On native Korean speakers' perception of moraic obstruents]. *Nihongo Kyōiku* 62. 179–193.

Min, Kwang Joon. 2007. The cause of the occurrence of geminates insertion: Evidence from Korean learners' production of the Japanese voiceless stops as geminates. *Journal of the Phonetic Society of Japan* 11(1). 58–70.

Minagawa, Yasuyo. 1994. Bogo kanshō sareta heisaon no muon jikan, VOT ni tsuite: Nanakakokugo kaku bogo washa no hatsuwa shiryō yori [On closure duration and VOT of stops affected by other languages: Production data by native speakers of 7 languages]. *Nihongo Kyōiku Gakkai Shūki Taikai Yokōshū*, 100–104.

Minagawa, Yasuyo. 1996. Sokuon no shikibetsu ni okeru akusento-gata to shiinshu no yōin: Kankoku-, tai-, chūgoku-, ei-, supein-go bogo washa no baai [Effects of pitch accent patterns and consonant types on the discrimination of moraic stops by native speakers of Korean, Thai, Chinese, English, and Spanish]. *Nihongo Kyōiku Gakkai Shunki Taikai Yokōshū*, 97–102.

Minagawa, Yasuyo. 1997. Chōon tan'on no shikibetsu ni okeru akusento-gata to onsetsu ichi no yōin: Kankoku-, tai-, chūgoku-, ei-, supein-go bogo washa no baai [Effects of pitch accent patterns and syllable positions on the discrimination of long and short vowels by native

speakers of Korean, Thai, Chinese, English, and Spanish]. *Nihongo Kyōiku Gakkai Shunki Taikai Yokōshū*, 123–128.
Minagawa, Yasuyo and Shigeru Kiritani. 1996. Discrimination of the single and geminate stop contrast in Japanese by five different language groups. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 30. 23–28.
Minagawa, Yasuyo, Kikuo Maekawa and Shigeru Kiritani. 2002. Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers. *Journal of the Phonetic Society of Japan* 6(2). 88–97.
Minagawa-Kawai, Yasuyo, Koichi Mori and Yutaka Sato. 2005. Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of Cognitive Neuroscience* 17(9). 1376–1385.
Mizutani, Osamu. 1976. Onsei kyōiku no mondaiten [Issues in phonetic education]. *Nihongo Kyōiku Kenkyū* 13. 1–7.
Motohashi-Saigo, Miki and Debra M. Hardison. 2009. Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning and Technology* 13(2). 29–47.
Najoan, Franky R., Noriko Yokoyama, Kazuhiro Isomura, Yo Usami and Yoshiko Kubota. 2012. The acquisition of Japanese vowel length contrast by Indonesian native speakers: Evidence from perception and production. *Journal of the Phonetic Society of Japan* 16(2). 28–39.
Nakagawa, Michiko and Toshiya Futamura. 2000. Shokyū nihongo gakushūsha no chōtan boin no ninshiki keikō to jizoku jikan [Auditory perception of long and short vowels by beginning learners of Japanese in relation to duration time]. *Hokkaidō Daigaku Ryūgakusei Sentā Kiyō* 4. 18–37.
Nakamichi, Makio. 1980. A study in rhythmic unit of Japanese. *Onsei Gengo no Kenkyū* 1. 15–22.
Nakata, Hitomi. 2002. Correlations between musical and Japanese phonetic aptitudes by native speakers of English. *Reading Working Papers in Linguistics* 6. 1–23.
Nishida, Fuminobu. 2003. Honkonjin nihongo gakushūsha no yūsei musei heisaon no sanshutsu to chikaku ni tsuite [On production and perception of Japanese voiced and voiceless stops by learners of Japanese in Hong Kong.]. *Nihongakkan* 7. 87–99.
Nishigori, Jiro. 1986. Gengoon no kategorī chikaku: Taiwankei nihongo gakushūsha no [t’] [t] [d] [r] no benbetsu o megutte [Categorical perception of speech sounds: On Taiwanese learners’ discrimination of [t’], [t], [d], and [r]]. *Nihongo to Nihongo Kyōiku* 15. 87–94.
Nishinuma, Yukihiro, Masako Arai and Takako Ayusawa. 1996. Perception of tonal accent by Americans learning Japanese. *Proceedings of ICSLP 96*, Philadelphia, 2, 646–649.
Nozawa, Motoko. 1974. Taiwanjin ryūgakusei no nihongo gakushū ni okeru onsei no shomondai ni tsuite no hōkoku [A report on phonetic issues in Taiwanese learners of Japanese]. *Nihongo to Nihongo Kyōiku* 3. 21–38.
Nozawa, Motoko and Jun Shigematsu. 1998. Kantongo washa no nihongo gakushū ni okeru akusento no mondai ni tsuite (1) [On the issue of Japanese pitch accent as learned by native speakers of Cantonese (1)]. *Nihongo to Nihongo Kyōiku* 27. 1–20.
Nozawa, Motoko and Jun Shigematsu. 2003. Inritsu to kanshō (3) [Prosody and interference (3)]. *Nihongo to Nihongo Kyōiku* 32. 107–130.
Nozawa, Motoko and Jun Shigematsu. 2006. Inritsu to kanshō (Musubi) [Prosody and interference (Conclusion)]. *Nihongo to Nihongo Kyōiku* 34. 83–108.
Oguma, Rie. 2000. Eigo bogo washa ni yoru chōon to tan’on no chikaku [Perception of long and short vowels by native speakers of English]. *Sekai no Nihongo Kyōiku* 10. 43–55.
Oguma, Rie. 2001. Nihongo gakushūsha ni yoru chōon to tan’on no sanshutsu [Production of Japanese short and long vowels by English-speaking learners]. *Takushoku Daigaku Nihongo Kiyō* 11. 79–87.

Oguma, Rie. 2002. Nihongo no chōon to tan'on ni kansuru chūkan gengo kenkyū no gaikan [A review of interlanguage studies in Japanese long vowels and short vowels]. *Gengo Bunka to Nihongo Kyōiku* 5. 189–200.

Ota, Mitsuhiko. this volume. Chapter 17: L1 phonology: phonological development. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Ota, Shinya. 2003. Chūgokugo bogo washa ni okeru sokuon no shidōhō ni tsuite no kōsatsu: VT-hō to jūrai no hatsuon shidō o hikaku shite [On the training methods for moraic obstruents for Chinese native speakers: Comparison of the VT method with extant pronunciation training methods]. *Takushoku Daigaku Gengo Kyōiku Kenkyū* 3. 79–86.

Otake, Takashi. this volume. Chapter 12: Mora and mora-timing. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton.

Otsubo, Kazuo. 1981. Nihonjin no sokuon no umu no handan ni tsuite [On the distinction between single and geminate obstruents in Japanese]. *Nagoya Daigaku Sōgō Gengo Sentā Gengo Bunka Ronshū* 3(1). 39–47.

Pisoni, David B. and Scott E. Lively. 1995. Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 433–459. Timonium, Maryland: York Press.

Roberge, Claude, Masayasu Kimura and Yoshikazu Kawaguchi. 1996. *Nihongo no hatsuon shidō: VT-hō no riron to jissai* [Pronunciation training for Japanese: Theory and practice of the VT method]. Tokyo: Bonjinsha.

Sadakata, Makiko and Kaoru Sekiyama. 2011. Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica* 138(1). 1–10.

Saita, Izumi, Kenichi Kido, Yoshiyuki Kawazoe, Hiroshi Kanai and Yasushi Inoguchi. 1992. Gaikokujin nihongo gakushūsha no tame no tokushu haku hatsuon shūtoku shien shisutemu [The system assisting the acquisition of special mora pronunciation for learners of Japanese as a foreign language]. In Shigeru Kiritani and Motohisa Imaishi (eds.), *Nihongo onsei no kenkyū to nihongo kyōiku*, 97–104. Monbushō Kagaku Kenkyūhi Hojokin Jūten Ryōiki Kenkyū "Nihongo Onsei" Kokusai Shinpojiumu Jikkō Iinkai.

Sakamoto, Emi. 2010. *An investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers*. Edinburgh: University of Edinburgh dissertation.

Sato, Mieko. 1986. Indoneshiajin gakushūsha ni taisuru nihongo no onsei kyōiku ni kansuru shōkō [On teaching pronunciation of Japanese to Indonesian learners]. *Nihongo Kyōiku* 60. 176–190.

Schachter, Jacquelyn. 1974. An error in error analysis. *Language Learning* 24. 205–214.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10(3). 209–231.

Shimizu, Katsumasa. 1999. A study on phonetic characteristics of voicing of stop consonants in Japanese and English. *Journal of the Phonetic Society of Japan* 3(2). 4–10.

Shport, Irina A. 2011. *Cross-linguistic perception and learning of Japanese lexical prosody by English listeners*. Portland, OR: University of Oregon dissertation.

So, Connie K. 2005. The influence of L1 prosodic background on the learning of Mandarin tones: Patterns of tonal confusion by Cantonese and Japanese naïve listeners. *Proceedings of the 2005 Annual Conference of the Canadian Linguistic Association*. http://westernlinguistics.ca/Publications/CLA-ACL/So.pdf (accessed 21 December 2012)

Sonu, Mee, Hiroaki Kato, Keiichi Tajima, Reiko Akahane-Yamada and Yoshinori Sagisaka. 2013. Nonnative perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics* 22. 373–398.

Sonu, Mee, Keiichi Tajima, Hiroaki Kato and Yoshinori Sagisaka. 2009. Perceptual training of singleton and geminate stops in Japanese language by Korean learners. *Proceedings of Interspeech 2009*, 1747–1750.

Sonu, Mee, Keiichi Tajima, Hiroaki Kato and Yoshinori Sagisaka. 2011a. Perceptual studies of Japanese geminate insertion phenomena based on timing control characteristics. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1886–1889.

Sonu, Mee, Keiichi Tajima, Hiroaki Kato and Yoshinori Sagisaka. 2011b. Perceptual training of vowel length contrast of Japanese by L2 listeners: Effects of an isolated word versus a word embedded in sentences. *Proceedings of Interspeech 2011*, 1137–1140.

Sonu, Mee, Keiichi Tajima, Hiroaki Kato and Yoshinori Sagisaka. 2012. Perceptual characteristics of Japanese geminate consonants by Korean native listeners: Focusing on the relationship between geminate insertion and Korean intervocalic tense consonants. *The Institute of Electronics, Information and Communication Engineers, Technical Report* 111(471). 7–12.

Sukegawa, Yasuhiko. 1999. Pitch realization of 2/3-mora-words by Brazilian learners of Japanese. *Journal of the Phonetic Society of Japan* 3(3). 13–25.

Sukegawa, Yasuhiko. 2001. What factors obstruct L2 phonology acquisition? *Journal of the Phonetic Society of Japan* 5(1). 75–76.

Sukegawa, Yasuhiko and Shigeru Sato. 1994. Kankokujin gakushūsha no nihongo akusento chikaku ni okeru onsetsu kōzō no eikyō [Effects of syllable structures on the perception of Japanese pitch accent by Korean learners]. *Tōhoku Daigaku Ryūgaku Sentā Kiyō* 2. 27–32.

Suzuki, Shinobu. 1963. Hatsuon no shidō to mondaiten [Teaching pronunciation and issues]. *Nihongo Kyōiku* 2. 7–20.

Tajima, Keiichi, Hiroaki Kato, Amanda Rothwell and Kevin G. Munhall. 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America* 123(1). 397–413.

Takakura, Tadahiro. 1988. Some practical suggestions for better pronunciation. *The Study of Sounds* (The Phonetic Society of Japan) 22. 71–79.

Toda, Takako. 1997. Strategies for producing mora timing by non-native speakers of Japanese. In D. G. Shūtoku Kenkyūkai (ed.), *Acquisition of Japanese as a second language*, 157–197. Tokyo: Bonjinsha.

Toda, Takako. 1998a. Nihongo gakushūsha ni yoru sokuon, chōon, hatsuon no chikaku hanchūka [Categorical perception of moraic obstruents, long vowels, and moraic nasal by learners of Japanese]. *Bungei Gengo Kenkyū* 33. 65–82.

Toda, Takako. 1998b. The mora and interlanguage syllable structure. *Journal of Japanese Language Teaching* 13. 23–45.

Toda, Takako. 2001. Hatsuon shidō ga akusento no chikaku ni ataeru eikyō [The effect of pronunciation practice upon the perception of Japanese accents]. *Bulletin of Center for Japanese Language, Waseda University* 14. 67–88.

Toda, Takako. 2003. Acquisition of special morae in Japanese as a second language. *Journal of the Phonetic Society of Japan* 7(2). 70–83.

Toda, Takako 2009. Nihongo kyōiku ni okeru gakushūsha onsei no kenkyū to onsei kyōiku jissen [Phonetic research and practice of teaching for Japanese language learners' pronunciation]. *Nihongo Kyōiku* 142. 47–57.

Toki, Satoshi. 1986. Onsei kyōiku no men kara mita kyōkasho [Teaching of pronunciation in textbooks]. *Nihongo Kyōiku* 59. 24–37.

Tsukada, Kimiko. 2011a. The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian. *Journal of the Acoustical Society of America* 129(2). 989–998.

Tsukada, Kimiko. 2011b. Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese. *Applied Psycholinguistics* 33(3). 501–516.

Tsurutani, Chiharu. 2003. Speech rate and the perception of geminate consonants and long vowels: A study of English-speaking learners of Japanese. *Nihongo Kyōiku* 119. 51–64.

Tsurutani, Chiharu. 2008. A computer program for pronunciation training and assessment in Japanese language classrooms: Experimental use of 'Pronunciation Check'. *Japanese Studies* 28(3). 305–315.

Ueyama, Motoko. 2012. *Prosodic transfer: An acoustics study of L2 English and L2 Japanese*. Bologna: Bononia University Press.

Wang, Nobuko. 1999. Chinese dialects which encourage acquisition of Japanese phonetic recognition for Chinese-speaking learners. *Journal of the Phonetic Society of Japan* 3(3). 36–42.

Watanabe, Shinichiro and Nobuo Hirato. 1985. Ni-onsetsugo ni okeru musei haretsuon to sokuon no handan kyōkai to senkō boin no nagasa no kankei [On the relationship between the perceptual boundary of voiceless plosives and their moraic counterparts and the duration of the preceding vowel in disyllables]. *Onsei Gengo* I. 1–8.

Wayland, Ratree P. and Bin Li. 2008. Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics* 36(2). 250–267.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.

Wu, Xian-Ghua, Jung-Yueh Tu and Yue Wang. 2011. Native and nonnative processing of Japanese pitch accent. *Applied Psycholinguistics* 33(3). 623–641.

Yamada, Tsuneo, Reiko A. Yamada and Winifred Strange. 1995. Perceptual learning of Japanese mora syllables by native speakers of American English: Effects of training stimulus sets and initial states. *Proceedings of the 13th International Congress of Phonetic Sciences* 1, 322–325.

Yamada, Yukihiro. 1963. Chōsenjin no nihongoon ninchi ni okeru nan'ido no sokutei ni tsuite [On the measurement of native Korean speakers' difficulty in perception of Japanese speech sounds]. *Nihongo Kyōiku* 3. 19–33.

Yamakawa, Kimiko and Shigeaki Amano. 2010. Perception boundary between fricative and affricate in Korean speakers. *Journal of the Acoustical Society of America* 128(4). 2352.

Yasuhara, Junko and Susumu Kasuya. 1994. The accent of double consonants in Japanese spoken by tone language speakers. *Bulletin of Kobe Women's University* 27(2). 165–173.

Yi, Kyong-Suk. 2003. The effect of articulation rate on Japanese double vowel production in Korean learners of Japanese. *Technical Report of the Institute of Electronics, Information and Communication Engineers of Japan* 103(307). 37–42.

Yokoyama, Kazuko. 1997. Ruikeironteki yūhyōsei to chōsengo washa no nihongo heisaon no yūsei musei no tairitsu no shūtoku [Typological markedness and the acquisition of stop voicing contrasts by native speakers of Korean]. *Gengo Bunka to Nihongo Kyōiku* 14. 41–51.

Yokoyama, Kazuko. 2000. Chūgokugo washa no nihongo heisaon shūtoku ni okeru konnan ten: Yūhyōsei to ruijisei no kanten kara [Native Chinese speakers' difficulty in acquiring Japanese plosives: From the perspectives of markedness and similarities]. *Tama Ryūgakusei Sentā Kyōiku Kenkyū Ronshū* 2. 1–11.

Ziolkowski, Michael and Karen Landahl. 1995. How successfully does visual feedback train listeners to produce and perceive non-native phonological contrasts? *Proceedings of the 13th International Congress of Phonetic Sciences* 1, 330–333.

Subject index